TOUCHPAD

11 CBSE

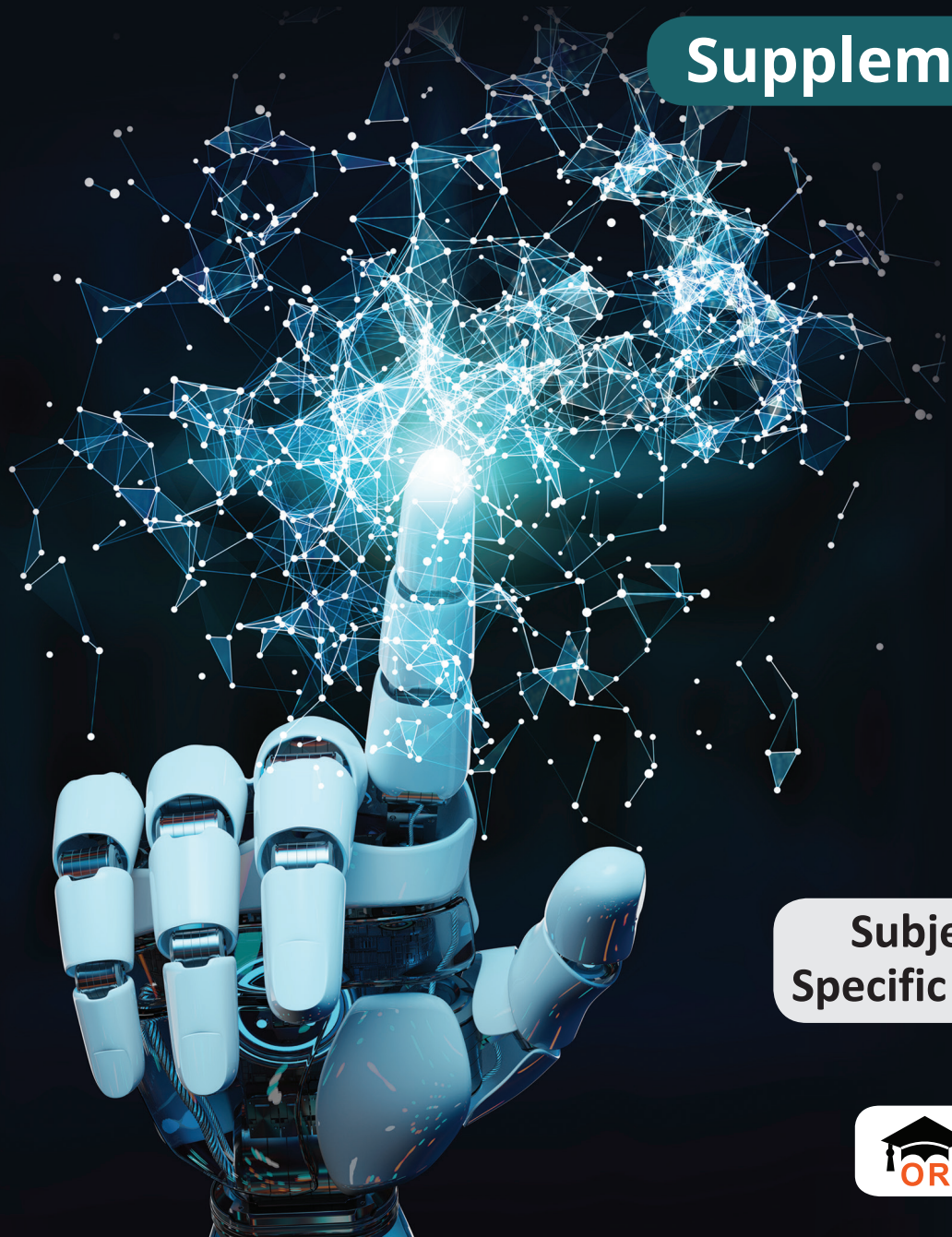
ARTIFICIAL INTELLIGENCE

CODE 843 | Skill Education

Based on 2024-25 Curriculum

MEENU KUMAR

Supplement



Subject Specific Skills

Ver. 2.0

ORANGE

ARTIFICIAL INTELLIGENCE (CODE 843)

CLASS-XI

Total Marks: 100 (Theory-50 + Practical-50)

OBJECTIVES OF THE COURSE

AI is a discipline in computer science that focuses on developing intelligent machines, machines that can learn and then teach themselves. These machines, then, can process vast amounts of data than humans can, and several times faster. However, AI can go across all disciplines to change the world for the better– from creating new healthcare solutions, to designing hospitals of the future, improving farming and our food supply, helping refugees acclimate to new environments, improving educational resources and access, and even cleaning our oceans, air, and water supply. The potential for humans to improve the world through AI is endless, as long as we know how to use it.

LEARNING OUTCOMES

In this course, the students will develop knowledge, skills and values to understand AI and its implications for our society and the world and to use AI to solve authentic problems, now and in the future. The students will engage with a host of multi-media online resources, as well as hands-on activities and sequence of learning experiences.

The following are the main objectives of the course:

1. Develop informed citizens with an understanding of AI and the skills to think critically and knowledgeably about the implications of AI for society and the world.
2. Develop engaged citizens with a rigorous understanding of how AI can be harnessed to improve life and the world we live in.
3. Stimulate interest and prepare students for further study to take up careers as AI scientists and developers to solve complex real world problems.

SCHEME OF UNITS

This course is a planned sequence of instructions consisting of units meant for developing employability and vocational competencies of students of Class XI opting for skill subject along with other education subjects. The unit-wise distribution of hours and marks for class XI is as follows:

	UNITS	NO. OF HOURS		MAX. MARKS
PART A	**Employability skills**			
	Unit 1: Communication Skills – III	15		2
	Unit 2: Self-Management Skills – III	10		2
	Unit 3: ICT Skills – III	15		2
	Unit 4: Entrepreneurial Skills – III	10		2
	Unit 5: Green Skills – III	10		2
	TOTAL	**60**		**10**
PART B	**Subject specific skills**	**Theory**	**Practical**	
	Introduction: Artificial Intelligence for Everyone	4	10	4
	Unlocking your Future in AI	6	10	5
	Python Programming	10	20	5
	Introduction to Capstone Project	6	15	5
	Data Literacy – Data Collection to Data Analysis	6	15	6
	Machine Learning Algorithms	9	15	6
	Leveraging Linguistics and Computer Science	5	10	5
	AI Ethics and Values	4	5	4
	TOTAL	**50**	**100**	**40**
PART C	**PRACTICAL WORK / PROJECT WORK**			
	IBM Skills Build Certification/any other industry certification			5
	Capstone Project			12
	Bootcamps/ Internship/other startups			7
	Practical File			10
	Written Exam (based on practical file)			10
	Viva Voce (based on practical file)			6
	TOTAL			**50**
	GRAND TOTAL			**100**

PART A: EMPLOYABILITY SKILLS

S. No.	Units	Duration in Hours
1.	Unit 1: Communication Skills – III	15
2.	Unit 2: Self-Management Skills – III	10
3.	Unit 3: Basic Information and Communication Technology Skills – III	15
4.	Unit 4: Entrepreneurial Skills – III	10
5.	Unit 5: Green Skills – III	10
	TOTAL	**60**

NOTE: Detailed Curriculum/ Topics to be covered under Part A: Employability Skills can be downloaded from CBSE website.

PART-B – SUBJECT SPECIFIC SKILLS

- Unit 1 – Introduction: Artificial Intelligence for Everyone
- Unit 2 – Unlocking your Future in AI
- Unit 3 – Python Programming
- Unit 4 – Introduction to Capstone Project
- Unit 5 – Data Literacy – Data Collection to Data Analysis
- Unit 6 – Machine Learning Algorithms
- Unit 7 – Leveraging Linguistics and Computer Science
- Unit 8 – AI Ethics and Values

UNIT 1 - INTRODUCTION: ARTIFICIAL INTELLIGENCE FOR EVERYONE

LEARNING OUTCOMES	THEORY	PRACTICAL
Students will be able to: • Communicate effectively about AI concepts and applications in written and oral formats. • Describe the historical development of AI. • Differentiate between various types and domains of AI, including their applications. • Recognize the key terminologies and concepts related to machine learning and deep learning. • Formulate informed opinions on the potential benefits and limitations of AI in various contexts.	• What is Artificial Intelligence? • Evolution of AI • Types of AI • Domains of AI • AI Terminologies • Benefits and limitations of AI	• Categorize the given applications into the three domains IBM Skills Build – Introduction to AI

UNIT 2 - UNLOCKING YOUR FUTURE IN AI

LEARNING OUTCOMES	THEORY	PRACTICAL
Students will be able to: • Articulate the demand for AI professionals and the diverse career opportunities available in the field. • Identify the requisite skills and tools needed to pursue a career in artificial intelligence. • Understand the potential roles and responsibilities of AI professionals across different industries. • Explore resources for further learning and skill development in the field of AI. • Evaluate their own interests and skills to determine potential pathways for a career in AI.	• The Global Demand • Some Common Job Roles In AI • Essential Skills and Tools for Prospective AI Careers • Opportunities in AI across Various Industries	• Identify ten companies currently hiring employees for in specific AI positions. • Note down the technical skills and soft skills listed by any two companies for the specific AI position. IBM Skills Build : Your Future in AI: The Job Landscape

UNIT 3 - PYTHON PROGRAMMING

LEARNING OUTCOMES	THEORY	PRACTICAL
Students will be able to: • Explain the basics of python programming language and write programs with basic concepts of tokens. • Use selective and iterative statements effectively. • Gains practical knowledge on how to use the libraries efficiently.	**Level 1 :** Basics of python programming, character sets, tokens, modes, operators, datatypes, Control Statements **Level 2 :** CSV Files, Libraries – Numpy, Pandas, Scikit-learn	• Minimum five programs to be taught using operators, data types, control statements (**Level 1**) • Minimum 5 programs on Numpy, Pandas, Scikit- learn (**Level 2**) IBM SkillsBuild - Python for Data Science

UNIT 4 - INTRODUCTION TO CAPSTONE PROJECT

LEARNING OUTCOMES	THEORY	PRACTICAL
Students will be able to: • Decompose any problem using the 5W1H method. • Apply Design thinking methodology. • Create empathy maps. • Align problems to SDGs. • Apply all the learnings in solving real world problems. • Comfortably express their solution to a problem in non-technical words.	• Design Thinking • Empathy Map • Sustainable Development Goals • Capstone Project	• Create an empathy map for a given scenario • Project Abstract Creation Using Design Thinking Framework IBM SkillsBuild - What is Design thinking?

UNIT 5 - DATA LITERACY – DATA COLLECTION TO DATA ANALYSIS

LEARNING OUTCOMES	THEORY	PRACTICAL
Students will be able to: • Explain the importance of data literacy in AI. • Identify different data collection methods and their applications. • Comprehend mathematical concepts related to matrices, its operations, and applications. • Apply basic data analysis techniques to analyse data. • Visualize the data using different techniques.	• What is Data Literacy? • Data Collection • Exploring Data • Statistical Analysis of data • Representation of data, Python Programs for Statistical Analysis and Data Visualization • Introduction to Matrices • Data Pre-processing • Data in Modelling and Evaluation	• Identification of the level of measurement • Python programs to demonstrate the use of mean, median, mode, standard deviation and variance • Python programs to visualise the line graph, bar graph, histogram, scatter graph and pie chart using matplotlib IBM SkillsBuild- Data Visualisation with Python (Modules 1,2,3)

UNIT 6 - MACHINE LEARNING ALGORITHMS

LEARNING OUTCOMES	THEORY	PRACTICAL
Students will be able to: • Differentiate the different types of machine learning methods. • They will be able to understand the concept behind each machine learning methods. • Apply these methods to develop simple solutions for some day-to- day situations. • Build up this knowledge to the next level to apply during Capstone Project development.	• Machine Learning in a nutshell • Types of Machine Learning • Supervised Learning • Understanding Correlation, Regression, Finding the line, Linear Regression algorithm • Classification – How it works, Types, k – Nearest Neighbour algorithm • Unsupervised Learning • Clustering – How it works, Types, k -means Clustering algorithm	• Calculation of pearson correlation coefficient in MS – Excel. • Demonstration of Linear regression in MS – Excel / using python program. • Demonstration of k – Nearest Neighbour using python program. • Demonstration of k – means clustering using python program. IBM SkillsBuild - Machine learning with Python

UNIT 7 - LEVERAGING LINGUISTICS AND COMPUTER SCIENCE

LEARNING OUTCOMES	THEORY	PRACTICAL
Students will be able to: • Develop a better understanding of the complexities of language and the challenges involved in NLP tasks. • Learn new techniques and algorithms for NLP tasks.	• Understanding Human Language Complexity • Introduction to Natural Language Processing (NLP) - Emotion Detection and Sentiment Analysis, Classification Problems, Chatbot • Phases of NLP • Applications of NLP	• Write an article on "IBM Project Debater – Interesting facts" • Create a chatbot on ordering ice-creams using any of the following platforms: o Google Dialogflow o Botsify.com o Botpress.com IBM SKillsBuild - Natural Language Processing

UNIT 8 - AI ETHICS AND VALUES

LEARNING OUTCOMES	THEORY	PRACTICAL
Students will be able to: • Demonstrate an understanding of the fundamental principles of ethics and gain insight into ethical considerations related to AI technologies. • Develop an understanding of AI bias, its sources, and its real-world implications, as well as the ethical considerations. • Identify and apply strategies for mitigating bias in AI systems to promote fairness and transparency in technology. • Recognize the significance of AI policies in promoting responsible, safe, and ethical use of AI technologies.	• Ethics in Artificial Intelligence • The five pillars of AI Ethics • Bias, Bias Awareness, Sources of Bias • Mitigating Bias in AI Systems • Developing AI Policies • Moral Machine Game • Survival of the Best Fit Game	• Summarize your insights and interpretations from the video "Humans need not apply." • Activity: Role Play on biased AI systems IBM SkillsBuild - AI Ethics • Comparative study of AI policies (that involve examining guidelines and principles) established by various organizations and regulatory bodies • Understanding ethical dilemma using • Moral machine Survival of the best fit

CONTENTS

Part B Subject Specific Skills

UNIT-2

UNLOCKING YOUR FUTURE IN AI

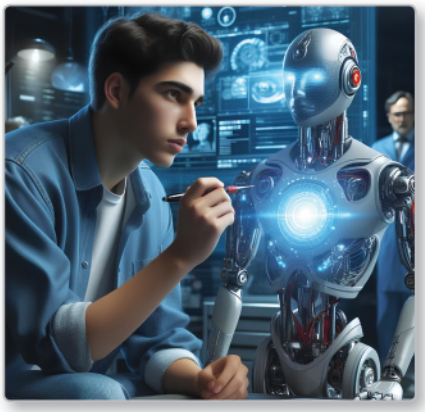
Learning Outcomes

- The Global Demand
- Essential Skills for Prospective Careers in AI
- Opportunities in AI across Various Industries
- Common Job Roles in AI
- Traits Required for an AI Professional
- Additional Learning Resources

Imagine a world where Artificial Intelligence (AI) powers advanced medical diagnostics and financial forecasting, enhancing your smartphone's predictive abilities, navigates self-driving cars, and manages virtual assistants seamlessly. This reality is shaped by AI, which is revolutionising industries and transforming how we live and work.

AI's influence spans from personalised recommendations to social media algorithms, offering endless possibilities for career paths. Whether in AI engineering or ethics, the future job landscape is rich with opportunities.

In this chapter, we explore AI's potential, unraveling its mysteries and uncovering exciting career avenues. Embracing AI as an opportunity for progress and creativity is essential in a world where its demand is soaring across sectors like healthcare, finance, and retail. Pursuing a career in AI offers diverse specialties like machine learning and robotics, fostering expertise development. As AI evolves, new job roles emerge, creating avenues for specialisation and growth.



Opportunities within the AI field have the potential to propel individuals into leading markets and lucrative income streams, particularly in sectors such as financial services, weather prediction, healthcare, and advanced robotics. Throughout this unit, we will delve into various job roles that, with the appropriate education and training, align with your skills and career aspirations.

The Global Demand

Once confined to the realms of science fiction, AI has seamlessly integrated into our daily routines. While contemplating career paths, it is worth exploring the abundant opportunities within the AI sector.

Despite concerns over potential job displacement by AI-driven automation, the encouraging reality is the proliferation of new employment avenues alongside the advancement of AI technology. The AI domain encompasses a wide spectrum of prospects, spanning machine learning, expert systems, data analysis, autonomous vehicles, and beyond.

Various industries are actively seeking AI integration to enhance their operations, including:

- Financial services
- Healthcare
- Technology sector
- Media industry
- Marketing field
- Retail sector
- Government agencies and military establishments
- National security initiatives
- Internet of Things (IoT) enabled systems
- Agricultural sector
- Gaming industry
- Weather forecasting and analysis
- E-commerce
- Telecommunications

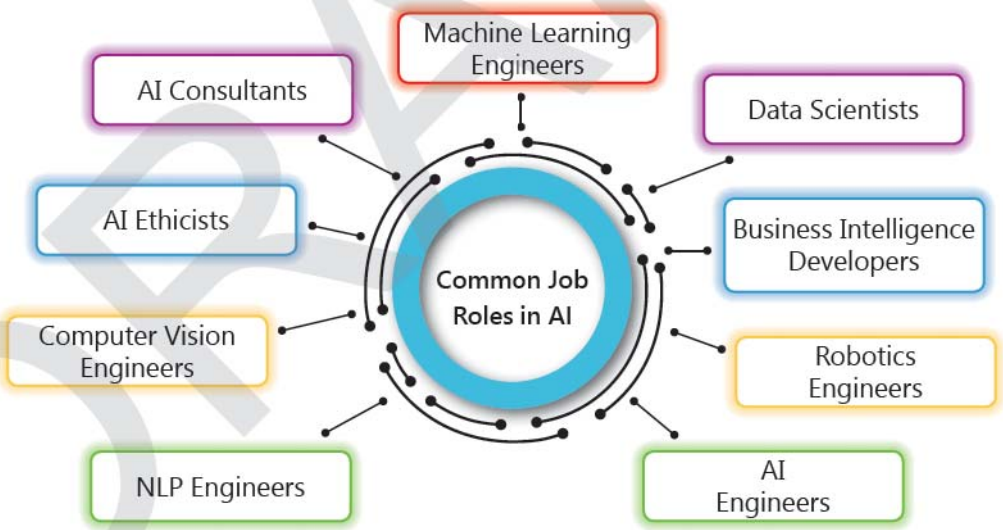
Embracing AI, presents a pool of career possibilities across diverse sectors, underscoring its significance in shaping the future landscape of numerous industries.

Brainy Fact

According to the World Economic Forum's Jobs Report 2023, the most significant developing career roles include AI and machine learning specialists, data analysts and scientists, and digital transformation specialists.

Common Job Roles in AI

Today's market has a wide range of high-demand career roles in the field of artificial intelligence (AI). Some jobs in AI include:

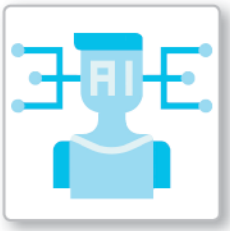


- **Machine Learning Engineers** combine software engineering and data science, using big data tools and programming frameworks to create scalable models for real-time data. Their responsibilities include conducting research, designing, and building artificial intelligence systems that are responsible for machine learning, as well as maintaining and improving the artificial intelligence systems that are currently in use as per company's perspective. Strong mathematical skills, familiarity with machine learning and deep learning, and proficiency in programming languages such as Java, Python, and Scala are required for success in this position.

- **Data Scientists** use machine learning and predictive analytics to get insights from massive datasets to take impactful business decisions. They are responsible for collecting, analysing, and interpreting data in decision-making. The data scientist role combines elements of different technical jobs, including mathematician, scientist, statistician, and computer programmer. This job profile requires proficiency in big data platforms such as Hadoop, Pig, and Spark, along with fluency in programming languages such as SQL, Python, and Scala, and a thorough understanding of descriptive and inferential statistics.

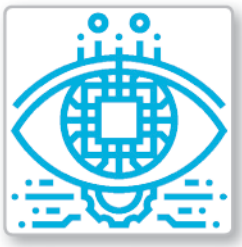
- **Business Intelligence Developers** analyse corporate and market trends to improve profitability and efficiency. Their job profile includes creating and managing business intelligence (BI) solutions, making technical questions, generating accurate search requests, turning data into easy-to-understand business formats, working with business analysts and colleagues to handle data, making visual tools to show data, and ensuring data is securely stored and backed up. This post requires strong technical and analytical abilities, as well as competence in data warehouse design and business intelligence technology.

- **Robotics Engineers** construct and manage AI-powered robots and mechanical devices that can perform tasks based on human commands. Their tasks involve researching different areas within robotics, such as nanotechnology, creating processes and prototypes for machine construction, and conducting tests on robotic systems. Programming proficiency, as well as knowledge of fields like mechanical engineering and electrical engineering are required for success in this field.

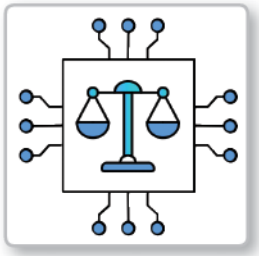


- **AI Engineers** develop and maintain AI applications using cutting-edge technologies. Their task involves developing diverse AI applications, ranging from contextual advertising, utilising sentiment analysis to visual identification or perception and language translation. Proficiency in software engineering, programming languages, and statistical analysis, along with strong knowledge of computer science, engineering, or related subjects is required in this field.

- **Natural Language Processing (NLP) Engineers** focus on voice assistants, speech recognition, and document processing. Their task involves reviewing and refining data science prototypes, crafting NLP applications, choosing suitable annotated datasets for Supervised Learning approaches, employing efficient text representations to convert natural language into valuable features, identifying and integrating appropriate algorithms and tools for NLP objectives. A person with speciality in computational linguistics, or a combination of computer science, mathematics, and statistics, is usually required in this field.



- **Computer Vision Engineers** create algorithms and systems to analyse and interpret visual information in photos and movies. Computer vision engineers utilise research in computer vision and collaborate with object-oriented software to manage the processing and analysis of extensive datasets, aiming to facilitate predictive decision-making automation through visual cues like images and videos. Their expertise rests in developing software solutions that can analyse and process visual data, with knowledge of image processing techniques as well as programming languages like Python and C++.



- **AI Ethicists** oversee the ethical aspects of AI technology development and implementation, assuring responsible and ethical use. Their main duty includes creating ethical guidelines, auditing AI systems for biases, collaborating with stakeholders, staying updated on AI advancements, engaging with the public, advising policymakers, assessing ethical risks, and partnering across disciplines. They advise on ethical frameworks, rules, and practices to promote fairness, transparency, and accountability in AI systems, which require a background in ethics, philosophy, or law, as well as competence in AI technology.

- **AI Consultants** provide expert counsel and advice to organisations on how to use AI technologies to address business problems and promote innovation. Their task involves decision-making, overseeing processes from planning to validation. They guide initiatives, assess existing infrastructure, and ensure alignment with business goals. Through rigorous evaluation, they determine optimal AI solutions, while post-deployment monitoring that guarantees sustained performance and reliability. They evaluate business needs, identify opportunities for AI integration, and create strategic AI initiatives. For this a thorough understanding of AI technologies, business processes, and industry trends, as well as effective communication and analytical abilities is required.

Ai Task

Match the job roles with the respective responsibilities. Research more about each job role to gain a deeper understanding.

Job Roles	Descriptions/Responsibilities
1. AI Engineer	a. Analyses and interprets complex data to derive insights.
2. Data Scientist	b. Designs and builds robotic systems with AI capabilities.
3. AI Ethicist	c. Specialises in processing and understanding human languages using AI techniques.
4. Machine Learning Engineer	d. Designs and develops algorithms for AI systems.
5. Robotics Engineer	e. Ensures ethical use and responsible development of AI technologies.
6. Natural Language Processing (NLP) Engineer	f. Creates and trains models for machine learning applications.

Essential Skills for Prospective Careers in AI

A successful career in Artificial Intelligence necessitates a wide range of skills that includes technical skills , soft skills and baseline skills.

Technical Skills

Technical skills refer to the specific abilities and knowledge required to perform tasks related to a particular field or profession. In the context of Artificial Intelligence (AI), technical skills include a range of competencies necessary to design, develop, and implement AI solutions. Here are some key technical skills relevant to AI:

- **Programming languages (Python, R, Java, C++):** Versatile in key programming languages is essential for AI development. This enables creation of efficient and scalable solutions.

- **Frameworks and libraries (TensorFlow, SciPy, NumPy):** Experienced with leading AI frameworks and libraries. This facilitates implementation of advanced algorithms and data processing.
- **Neural networks:** Hands on experience in designing and training neural networks to develop models for tasks like image and speech recognition.
- **Machine learning:** Proficient in building and deploying machine learning models. Applies techniques to extract valuable insights from data.
- **Deep learning:** Expert in deep learning methodologies. Constructs complex models for sophisticated AI applications.
- **Shell scripting:** Adept at writing shell scripts to automate tasks. Enhances system efficiency and workflow management.
- **Cluster analysis:** Skilled in performing cluster analysis to identify patterns. Segments data for detailed and actionable insights.
- **Visualisation of data:** Hands on experience in using data visualisation tools like Tableau, Microsoft Power BI, etc. for creating interactive dashboards to support decision-making. Develops comprehensive reports and visualisations for strategic insights.
- **Knowledge of sensor fusion**: For integrating data from multiple sensors (cameras, LiDAR, and so on) to create comprehensive visual understanding systems.

Soft Skills/Workplace Skills

Soft skills, also known as interpersonal skills or people skills, are non-technical skills that are essential for successful interaction and communication with others, both individually and in groups or teams. These skills are important in every profession and are particularly crucial in roles that involve frequent interaction with clients, customers, colleagues, or the public. Here are some key soft skills:

- **Communication skills:** Effective communication skills are essential for communicating complicated technological concepts to non-technical stakeholders and working with multidisciplinary teams.
- **Teamwork and collaboration:** Effective cross-functional teams require strong teamwork and collaboration skills to produce AI solutions.
- **Problem solving:** AI requires the ability to identify, analyse, and resolve challenges effectively.
- **Decision making:** The process of selecting the best course of action from varied options results in best AI resolve.
- **Analytical thinking:** AI initiates analytical thinking skills to examine and interpret complex information to make informed decisions.
- **Time management:** Efficiently organising and prioritising tasks to maximise productivity and meet deadlines.
- **Business intelligence:** Utilising data analysis and insights to drive informed decision-making and strategy development in business.
- **Critical thinking:** The ability to evaluate information objectively and make reasoned judgments or decisions.

Baseline Skills

Baseline skills refer to fundamental competencies that are essential for understanding, working with, and applying AI technologies effectively. These skills provide a foundation upon which individuals can build more advanced capabilities in AI development, implementation, and research. Here are some key baseline skills in AI:

- **Linear Algebra:** Mathematics dealing with vectors, matrices, and linear transformations, widely used in fields like computer graphics and machine learning.
- **Probability:** Branch of mathematics analysing random phenomena and likelihood of different outcomes, crucial for risk assessment, and statistical inference.

- **Statistics:** Science of collecting, analysing, and interpreting data, used in science, engineering, and social sciences for making data-driven decisions.
- **Signal Processing:** Analysis and interpretation of signals, such as audio and video data, used in telecommunications, medical imaging, and control systems.
- **Big Data:** Management and analysis of large and complex datasets, facilitated by technologies like Hadoop and Spark, applied across industries for extracting insights and value from data.

Your Professional Toolkit

In addition to learning the requisite skills, AI practitioners should become familiar with prominent AI tools, platforms, and programming languages. Few important tools and frameworks are as follows:

- **Python** is a versatile programming language with pre-built libraries for advanced and scientific computing.
- **R** is a computer language used for data collecting, organising, and analysing, with applications in machine learning and statistics.

- **Java** is a commonly used language in AI to develop intelligent programs, neural networks, and machine learning solutions.
- **C++** is flexible and object-oriented language, making it ideal for procedural programming and hardware manipulation in AI.
- **TensorFlow** is an open-source machine learning framework that provides tools and libraries for building advanced AI applications.
- **SciPy** is an open-source Python library, used for solving scientific and mathematical problems. It helps users to manipulate and visualise data using various commands.
- **NumPy** is a Python-based package used for scientific computing and advanced mathematical operations while managing massive data set.

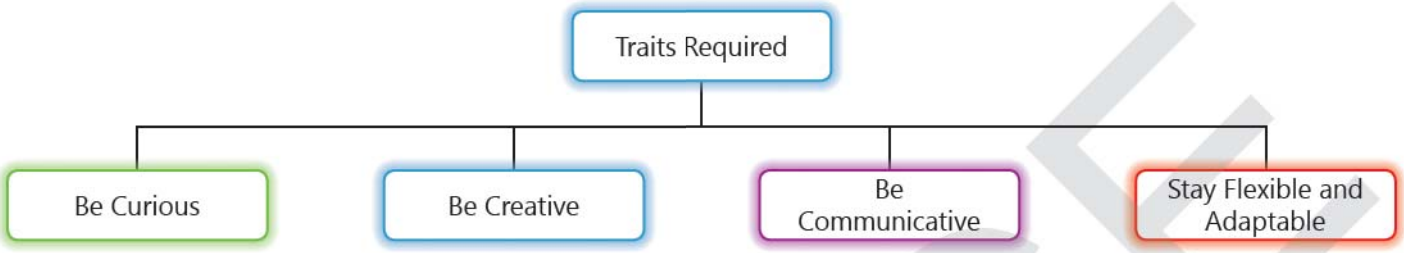
Aspiring AI professionals can prepare themselves for success in this dynamic and quickly growing sector by learning the necessary technical skills and tools. Whether you want to develop AI algorithms or apply AI solutions in real-world applications, establishing a solid foundation of skills, and expertise is essential for pursuing fascinating career options in artificial intelligence.

Traits Required for an AI Professional

If you are an aspiring AI professional, it's useful to have the following general characteristics:

- **Be curious:** Curiosity is essential for comprehending the business problems, analysing the data, and seeking potential data applications.

- **Be creative:** It is important to think about how AI could tackle various issues. Creativity helps you view problems from different angles and envision solutions others might overlook.
- **Be communicative:** It is crucial to effectively convey your ideas. While discovering solutions can be thrilling, you must also be able to explain their significance and the opportunities they present to others.
- **Stay flexible and adaptable:** AI projects often yield unexpected results. Embrace continuous changes and be prepared to learn from your experiences, making necessary adjustments even if it means starting from nothing.



Brainy Fact

"The skills that matter the most in the era of AI are not just AI skills, but also human and soft skills. These are the skills that help us work with others, solve problems, and think critically. The top five skills that employers in India want the most are problem-solving abilities, AI skills, critical thinking, communication skills, and IT and web skills."

— stated by Tomer Cohen, chief product officer at LinkedIn

Experiential Learning

Video Session

Scan the QR code or visit the following link to watch the video: Machine Learning Roadmap 2024 | Machine Learning Career Path 2024

https://www.youtube.com/watch?v=JVg8d2smoiQ

After watching the video answer the following questions:

1. Identify three AI tools that transforms natural language prompts into coding suggestions streamlining the programming process.

2. Name five companies, following the path of AI using Machine Learning.

AI Reboot

1. Fill in the blanks:
   a. ..... Engineers create algorithms and systems to analyse and interpret visual information in photos and movies.
   b. Two technical skills that an AI engineer must have are ..... and ..... .
   c. AI Consultants provide expert advice to firms on how to use AI technologies to address business problems and promote ..... .
2. Answer the following:
   a. List two general traits that an AI professional should have.
   b. Imagine you are a Data Scientist.
      i. List two programming language that your professional toolkit must have.
      ii. Which two soft skills are most important for you in the workplace?

Opportunities in AI across Various Industries

Artificial Intelligence specialists develop and implement AI systems that leverage machine learning and neural networks to forecast trends, improve consumer experiences and suggestions, and deliver answers to complex problems. While few AI specialists strive towards the aim of General AI (interconnected systems that can be as creative as humans), others concentrate on specific applications.

The table below provides you with a selection of options based on your subject choices:

Industry	Employment opportunities	Job/Position	Job role	Subject knowledge required
Agriculture	Monitoring crop health, managing irrigation, and increasing yields.	Precision Agriculture Specialist	Develops AI-powered drones and sensors to monitor crops.	Biology, Mathematics, Computer Science/ Artificial Intelligence
		Crop Yield Prediction Analyst	Predicts agricultural yields using AI methods.	
		Livestock Monitoring Specialist	Monitors the health and productivity of agricultural animals.	
Automobile	Design, manufacturing, and motor vehicles sales.	Autonomous Vehicle Engineer	Creates AI algorithms for self-driving vehicles.	Mathematics, Physics, Computer Science/Artificial Intelligence
		Simulation Engineer	Develops virtual settings to assess autonomous car technologies.	
		Robotics Engineer	Develops AI-powered robots for automotive applications.	

Industry	Employment opportunities	Job/Position	Job role	Subject knowledge required
Banking	Loan approval automation, fraud detection, and customised financial guidance.	Loan Approval Specialist	Uses AI to automate the loan approval procedure.	Economics, Mathematics, Computer Science/ Artificial Intelligence
		Fraud Detection Analyst	Makes use of AI algorithms to detect fraudulent transactions.	
		Financial Advisor	Uses AI-driven data to provide individualised financial advice.	
Beauty & Wellness	Skincare analysis, virtual styling, and health tips.	AI-powered Skincare Assistant	Offers personalised skincare advice.	Chemistry, Biology, Computer Science/ Artificial Intelligence
		Virtual Hair Stylist	Simulates different hairstyles using AI.	
		Wellness Chatbot	Offers guidance on nutrition and fitness.	
Design	Content production, design optimisation, and user experience enhancement.	Generative Design Assistant	Uses AI techniques to optimise design solutions.	Art & Design, Computer Science/Artificial Intelligence, Mathematics
		AI-powered UX Designer	Uses AI-driven insights to improve user experience.	
		AI-powered Content Creator	Uses AI technologies to create content.	
Fashion	Trend research, virtual try-on, and personalised fashion advice.	AI-powered Fashion Stylist	Makes customised outfit suggestions with AI.	Fashion Design, Mathematics, Computer Science/ Artificial Intelligence
		Trend Analyst	Uses AI algorithms to analyse current trends in fashion industry.	
		Virtual Clothes Try-on Specialist	Uses AI and Augmented Reality to enable virtual clothes try-on.	
Finance	Risk management, fraud detection, market analysis, and recommendations for investments.	Quantitative Analyst	Uses AI algorithms to analyse market patterns.	Economics, Mathematics, Computer Science/ Artificial Intelligence
		Fraud Detection Analyst	Makes use of AI models to find fraudulent activity.	
		Financial Advisor	Gives individualised financial advice based on insights powered by AI.	

Industry	Employment opportunities	Job/Position	Job role	Subject knowledge required
Geospatial	Remote sensing, mapping, and spatial data analysis technologies.	Geographic Information Systems (GIS) Specialist	Uses AI to analyse spatial data.	Geography, Geology, Computer Science/Artificial Intelligence
		Remote Sensing Analyst	Uses AI algorithms to interpret satellite pictures.	
		Mapping Technician	Maps using drones with AI capabilities.	
Government & Military	Predictive analytics, Surveillance, citizen services, and military technologies.	National Security Analyst	Utilises AI-powered surveillance technologies.	Political Science, Computer Science/Artificial Intelligence, Mathematics
		Defence Contractor	Creates AI-enabled military technology.	
		Government AI Specialist	Uses AI to improve citizen services and comply with regulations.	
Healthcare	Drug discovery, medical imaging analysis, and personalised medical care.	Medical Imaging Analyst	Uses AI algorithms to analyse medical images.	Biology, Chemistry, Computer Science/ Artificial Intelligence
		Virtual Nurse Assistant	Offers personalised health advice.	
		Drug Discovery Researcher	Uses AI to find potential drug prospects.	
Information Technology	Create AI algorithms, systems, and infrastructure for a variety of applications.	Machine Learning Engineer	Builds AI systems and algorithms.	Computer Science/Artificial Intelligence, Mathematics, Physics
		AI Developer	Develops AI-enabled apps.	
		AI Infrastructure Specialist	Manages and optimises AI infrastructure.	
Media	Creating visual effects, developing content, and analysing audience behaviour.	Visual Effects Artist	Produces visual effects with AI tools.	Fine Arts, Media Studies, Computer Science/Artificial Intelligence
		Content Creator	Produces material with insights from AI.	
		Audience Analyst	Employs AI algorithms to examine audience behaviour.	

Industry	Employment opportunities	Job/Position	Job role	Subject knowledge required
Retail	Optimising inventories, predicting sales, and improving customer experience.	Inventory Management Specialist	Uses AI techniques to optimise inventory levels.	Business Studies, Mathematics, Computer Science/ Artificial Intelligence
		Sales Forecasting Analyst	Predicts sales using AI models.	
		Customer Experience Designer	Improves the customer experience through AI-driven insights.	
Sales & Marketing	Sales forecasting, customer segmentation, and campaign automation.	Marketing Campaign Automation Specialist	Uses AI to automate marketing campaigns.	Business Studies, Mathematics, Computer Science/ Artificial Intelligence
		Customer Segmentation Analyst	Uses AI to segment clients according to their behaviour.	
		Sales Forecasting Analyst	Uses AI models to forecast sales.	
Textile	Inventory management, quality assurance, and fabric design.	AI-powered Fabric Design Specialist	Utilising AI to design inventive textile patterns.	Chemistry, Art & Design, Computer Science/Artificial Intelligence, Mathemtics or Accountancy
		Textile Quality Control Inspector	Uses AI-enabled systems to ensure product quality.	
		Smart Inventory Management Specialist	Uses AI algorithms to optimise inventory levels.	
Tourism	Personalised travel suggestions, client service, and itinerary planning.	Travel Recommendation Engine Developer	Offers personalised travel recommendations	Geography, Business Studies, Computer Science/ Artificial Intelligence
		Chatbot for Customer Service	Helps travellers with bookings and inquiries.	
		Smart Travel Itinerary Planner	Optimises travel routes and schedules.	

Brainy Fact

Glassdoor (a free website where current and past employees may anonymously review employers, post photos, and share other company-related content) ranked the 50 Best Jobs in America for 2022 based on earnings potential, work satisfaction, and job opportunities. AI-related jobs, such as data scientist and machine learning engineer, are placed third and sixth respectively.

Experiential Learning

Video Session

Scan the QR code or visit the following link to watch the video: 10 Ways AI is Transforming Industries Right Now

https://www.youtube.com/watch?v=rUI3r5inwz4

After watching the video answer the following questions:

1. Identify five ways in which AI is transforming industries.

2. Explain five ways in which AI in automation is making our lives easy and comfortable.

Ai Task

Generate music with the help of AI! Visit the website https://soundraw.io/

Click on the Start Customizing button. Now click on any music track from the list that appears. Make changes on the mixer as per your choice!

Next note down the skills you think you should have to create such a website to generate music using AI. Do you think you can do it?

Additional Learning Resources

For those interested in learning more about artificial intelligence and staying updated with the latest advancements in the subject, here are some resources you can browse, bookmark, and keep in mind. Depending on your interests, you can explore a wide range of organisations and websites.

Blogs and News to Remain Up to Date

Blogs and news sites are important platforms to stay updated. Here are some of the websites and platforms:

- **Analytics Insight** provides the most recent information, insights, trends, opinions, and a magazine with the thoughts and perspectives of leading business executives and industry leaders who discuss their successes, failures, and lessons learned in building successful companies.
- **KDnuggets** is a premier website on data science, machine learning, artificial intelligence, and analytics KD stands for Knowledge Discovery.
- **Towards Data Science** is an online publication where independent authors who abide by their rules and principles can publish their work, share their knowledge and skills, and engage a vast audience on medium. It offers top-notch instructional resources, webinars, online events, and courses.
- One of the best online resources for data practitioners is **Data Science Central**. Data Science Central offers a community experience that combines a comprehensive editorial platform, social engagement, forum-based support, and the latest information on technology, tools, trends, and professions. It covers everything from statistics and analytics to machine learning and AI.
- **Datanami** is a news platform devoted to providing analysis, insight, and the latest information about big data trends and solutions. The portal sheds light on all cutting edge technologies including networking, storage and applications, and their effect upon business, industry, government, and research.

Build Skills using These Free Websites

In today's environment, it is very important to keep yourself updated in terms of skills as well. Here is a list of some free websites where you can develop or build your skills for free.

- With the help of work skills, courses, digital certifications, and more, IBM SkillsBuild (https://skillsbuild.org/) can help you propel your career in technology.
- To assist you in developing the abilities required to work on independent data science projects, Kaggle (https://www.kaggle.com/) provides free online micro courses. By participating in Kaggle competitions, you can enhance your knowledge of data science and machine learning on Kaggle.
- Introduction To Artificial Intelligence is a brief, hands-on course that is one of many free video-based courses on Artificial Intelligence available on Udemy (https://www.udemy.com).
- Artificial Intelligence: Preparing Your Career for AI is another course that Udemy offers. It addresses the things you should be doing right now to get ready for AI's arrival.
- A summary of All the Math You Need to Know in Artificial Intelligence is available at https://www.freecodecamp.org/. An outline of the fundamental arithmetic concepts that are necessary to work in artificial intelligence is provided by Jason Dsouza.
- Understanding Machine Learning is a two-hour, free course offered by DataCamp (https://www.datacamp.com/) that introduces machine learning without the need for coding.
- The largest web developer website in the world, W3Schools (https://www.w3schools.com/), provides a range of free online tutorials with practical experience.
- Codecademy (https://www.codecademy.com) offers free coding classes on 12 various programming languages, including Python, Java, and C++. The website features tutorials on some prominent data science programming languages, like R, SQL, and Python.

More Details on Universities Providing Professional AI Courses

Seeing the advancement of AI in industries and its importance in developing one's career, universities have also upgraded themselves and are providing professional courses in AI and machine learning. Here are some of the universities that are offering courses on this:

- IIT Madras offers a four-year Bachelor of Science program in Data Science and Applications. (study.iitm.ac.in/ds)

- The All India Council for Technical Education's special "DigitalSkilling" webpage. Check out the many courses and internships available on this website. https://1crore.aicte-india.org/
- Nowadays, B.Tech programs in AI and ML, Data Science, Robotics, and Computer Science with specialisations are offered by the majority of India's best universities. Learn more about these courses, students can go to the websites of the colleges. A B.Sc in AI and ML is also offered by several universities.

Ai Task

Explore and write two AI roles, for which these companies are hiring.

1. Nvidia _____ _____
2. Salesforce _____ _____
3. Amazon _____ _____
4. Microsoft _____ _____

At a Glance

- AI powers smartphones, self-driving cars, and virtual assistants, revolutionising industries and daily life.
- AI offers diverse career paths, including engineering and ethics, across sectors like healthcare, finance, and retail.
- AI has integrated into our daily routines, with career opportunities in machine learning, data analysis, and autonomous vehicles.
- AI is sought after in financial services, healthcare, technology, media, marketing, retail, government, military, IoT, agriculture, gaming, and weather forecasting.
- Common AI job roles include AI Engineer, Data Scientist, AI Ethicist, Machine Learning Engineer, Robotics Engineer, and NLP Specialist.
- Key technical skills required for a career in AI include programming languages (Python, R, Java, C++), frameworks (TensorFlow, SciPy, Numpy), neural networks, machine learning, deep learning, shell scripting, cluster analysis, Tableau, and Microsoft Power BI.
- Important soft skills required for a career in AI are communication, teamwork, problem-solving, decision-making, analytical thinking, time management, business intelligence, and critical thinking.
- Curiosity, creativity, adaptability, and effective communication are important traits for success in AI.
- AI specialists work in fields like automotive, agriculture, retail, media, IT, healthcare, finance, government, tourism, beauty, banking, geospatial, textile, design, sales and marketing, telecommunication, E-commerce, and fashion.
- Tools like Soundraw.io can generate music, highlighting the need for programming and AI skills to create such platforms.
- AI-related jobs like data scientist and machine learning engineer rank high in job satisfaction, earning potential, and opportunities.
- Mastery of tools like TensorFlow and Python is essential for AI development.
- AI Ethicists ensure the responsible and ethical development of AI technologies.
- AI enhances predictive analytics, surveillance, and citizen services in government and military applications.
- AI is used for drug discovery, medical imaging analysis, and personalised medical care.
- The evolving AI field continually creates new job roles and specialisation opportunities, driving growth and innovation across industries.
- The integration of AI into everyday activities has increased demand for professionals skilled in areas like machine learning, data analysis, and autonomous vehicles.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. What is one of the main benefits of pursuing a career in AI?
   a. Limited job roles ◯ b. High demand across multiple industries ◯
   c. Requires no technical skills ◯ d. Low-income potential ◯
2. Which of the following is NOT a common job role in AI?
   a. AI Engineer ◯ b. Human Resource ◯
   c. Data Analyst ◯ d. Machine Learning Engineer ◯
3. Which industry is NOT actively seeking AI integration?
   a. Financial services ◯ b. Healthcare ◯
   c. Craft and Artisanal ◯ d. Agriculture ◯
4. Which of the following is a primary responsibility of a Machine Learning Engineer?
   a. Designing hardware components ◯ b. Building AI systems and algorithms ◯
   c. Managing IT infrastructure ◯ d. Writing user interface code ◯
5. What is the primary focus of AI Ethicists?
   a. Building AI models ◯ b. Designing robotic systems ◯
   c. Ensuring ethical use of AI technologies ◯ d. Analysing big data ◯
6. Which programming language is commonly used for AI development?
   a. HTML ◯ b. CSS ◯
   c. Python ◯ d. PHP ◯
7. Which soft skill is crucial for efficiently organising and prioritising tasks to maximise productivity and meet deadlines?
   a. Linear Algebra ◯ b. Time management ◯
   c. Deep learning ◯ d. Shell scripting ◯
8. What of the following is used for building advanced AI applications and is open-source?
   a. Java ◯ b. TensorFlow ◯
   c. Tableau ◯ d. Power BI ◯
9. In which sector are Autonomous Vehicle Engineers primarily employed?
   a. Retail ◯ b. Automobile ◯
   c. Media ◯ d. Tourism ◯
10. What is a significant challenge AI professionals must be prepared to handle in their projects?
    a. Stable and unchanging project requirements ◯ b. Predictable outcomes ◯
    c. Continuous change and unexpected results ◯ d. No need for adaptability ◯

B. Fill in the blanks.

1. AI's influence spans from personalised recommendations to _____ algorithms, offering endless possibilities for career paths.
2. To develop effective AI solutions, professionals need a combination of technical, soft skills, and _____skills.
3. A job that involves using statistical techniques to interpret complex data is called a _____ .
4. _____ involves the analysis and interpretation of signals, such as audio and video data, used in telecommunications and medical imaging.
5. A crucial skill for a _____ is expertise in image processing and recognition algorithms.

C. State whether the following statement is True or False.

1. Natural Language Processing (NLP) Engineers develop systems that can understand and generate human language. _____
2. AI specialists only work in the technology sector and have no roles in industries like healthcare or finance. _____
3. Precision Agriculture Specialist develops AI-powered drones and sensors to monitor crops. _____
4. AI Engineers are responsible for converting data into formats that are useful for analysis. _____
5. Trend Analyst uses AI algorithms to analyse current trends in fashion. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. What are the responsibilities of a Data Scientist in the AI field?

Ans. A Data Scientist analyses and interprets complex data, derives insights, builds predictive models, and helps in decision-making by providing data-driven solutions to business problems.

2. How do AI professionals use neural networks?

Ans. AI professionals use neural networks to design and train models for tasks such as image and speech recognition, enabling machines to mimic the human brain's ability to recognise patterns.

3. What soft skills are essential for AI professionals, and why?

Ans. Effective communication, teamwork, problem-solving, and critical thinking are essential soft skills for AI professionals as they enable collaboration, clear idea presentation, and innovative solutions to complex problems.

4. Describe the role of an AI Ethicist.

Ans. An AI Ethicist ensures the ethical use and responsible development of AI technologies, addressing issues like bias, fairness, transparency, and the societal impacts of AI systems.

5. Which industries benefit from AI in the context of customer experience improvement?

Ans. Retail and tourism industries benefit from AI through personalised recommendations, customer service chatbots, and optimised user experiences, enhancing customer satisfaction and loyalty.

B. Long answer type questions.

1. What are some dynamic career paths within AI that offer significant growth prospects across industries like healthcare, finance, and retail?

Ans. The demand for AI specialists is escalating as industries recognise its transformative potential. Job opportunities span diverse sectors like healthcare, finance, and retail. AI engineers design algorithms and systems, while data scientists extract insights from complex data. AI ethicists ensure responsible AI development. Robotics engineers create AI-powered machines, and NLP specialists process human languages. These roles offer dynamic career paths with significant growth prospects.

2. Explain the importance of soft skills for AI professionals and how they contribute to successful AI projects.

Ans. Soft skills are pivotal for AI professionals as they facilitate effective communication, collaboration, and problem-solving. Communication ensures clear understanding of project objectives, while teamwork fosters synergy within teams. Critical thinking aids in data analysis and algorithm optimisation. Creativity fuels innovation, enabling novel solutions to complex problems. These skills complement technical expertise, leading to successful AI projects that address real-world challenges.

3. What are the key responsibilities of a Machine Learning Engineer in the AI field?

Ans. Machine Learning Engineers develop, implement, and maintain machine learning models and systems. They preprocess data, select appropriate algorithms, and train models using large datasets. Engineers optimise models for accuracy and efficiency and deploy them into production environments. They also collaborate with cross-functional teams to integrate machine learning solutions into applications and ensure scalability and reliability.

4. Discuss the role of AI in the healthcare industry and its potential impact on patient care.

Ans. AI revolutionises healthcare by improving diagnostics, treatment planning, and patient outcomes. Medical image analysis aids in early disease detection, while predictive analytics enhances personalised treatment strategies. Virtual health assistants provide round-the-clock support, enhancing patient engagement and satisfaction. AI also streamlines administrative tasks, reduces errors, and lowers healthcare costs, leading to more efficient and accessible healthcare services.

5. Examine the significance of AI skills and soft skills in the context of future job paths and career growth.

Ans. AI skills like machine learning and deep learning are essential for technical roles, enabling professionals to develop advanced AI systems and algorithms. Soft skills such as communication, teamwork, and critical thinking are equally important, fostering collaboration and innovation. Professionals with a blend of technical and soft skills are well-equipped to navigate diverse job paths in AI, ensuring career growth and adaptability in a rapidly evolving field.

C. Competency-based/Application-based questions:

1. Sunita is a recent graduate who majored in computer science. She has always been fascinated by Artificial Intelligence (AI) and dreams of working in a role that involves developing algorithms for AI systems. After applying to several companies, she receives a job offer from a tech firm looking for someone to design and develop algorithms for their AI projects. Identify the role she might have got job for and write down the skills required for the same.

Ans. Elaboration: Sunita most likely received a job offer as a Machine Learning Engineer. This role involves designing and developing algorithms for AI systems, which aligns with Sunita's interest in Artificial Intelligence and her academic background in computer science.

Skills Required:

- Proficiency in programming languages such as Python, R, or Java.
- Strong understanding of machine learning concepts and algorithms.
- Knowledge of data structures and algorithms.
- Experience with frameworks like TensorFlow or PyTorch.
- Ability to work with large datasets and apply statistical techniques.
- Problem-solving skills to address complex AI challenges.
- Excellent communication skills to collaborate with team members and stakeholders.

2. Rajnish is working as a Natural Language Processing (NLP) Specialist at a major tech company. His job involves creating systems that can understand and process human language. One of his recent projects includes developing a chatbot that can interact with customers and provide them with support. What specific responsibility does Rajnish's job entail?

Ans. Rajnish's role as an NLP Specialist revolves around enabling machines to understand and process human language. His responsibilities include designing algorithms and models that can extract meaning from text, identify patterns, and generate contextually relevant responses. To excel in this role, Rajnish needs a deep understanding of linguistics, computational linguistics, and natural language processing techniques.

He must be proficient in programming languages like Python and have experience with NLP libraries and frameworks such as NLTK, SpaCy, or Transformers. Rajnish's expertise allows him to develop sophisticated NLP applications like chatbots, sentiment analysis tools, and language translation systems, enhancing user interactions and experiences.

3. Ankit is part of a cross-functional team working on a new AI product. He often collaborates with colleagues from different departments, including marketing, development, and design. Ankit excels at bringing everyone together to achieve a common goal, and his ability to work well with others is a key factor in the project's success. Which soft skill is Ankit demonstrating in this scenario?

Ans. Ankit demonstrates exemplary teamwork and collaboration skills by effectively bridging communication gaps between diverse teams working on the AI project. His ability to foster cooperation, resolve conflicts, and facilitate productive interactions enables the cross-functional team to work cohesively towards shared objectives. Ankit cultivates an inclusive and supportive environment where team members feel valued and motivated to contribute their expertise. His leadership qualities empower individuals to leverage their strengths and overcome challenges collectively, driving innovation and achieving project milestones. Ankit's adeptness at building relationships and fostering teamwork is instrumental in the project's success and organisational growth.

4. Deepali is an AI Ethicist at a prominent tech firm. Her role involves ensuring that the development and implementation of AI technologies are conducted responsibly and ethically. Recently, she was part of a team that reviewed a new facial recognition system to ensure it does not have biases against certain demographic groups. What is the primary responsibility of Deepali's job?

Ans. As an AI Ethicist, Deepali plays a pivotal role in ensuring that AI technologies are developed and deployed ethically and responsibly. Her primary responsibility involves evaluating the ethical implications of AI systems, identifying potential biases or risks, and advocating for fair and transparent practices. Deepali collaborates with cross-functional teams to integrate ethical considerations into the design, development, and implementation of AI technologies. She conducts thorough assessments and audits to mitigate harm and promote ethical decision-making throughout the AI lifecycle. Deepali's expertise in ethics, coupled with her technical acumen, helps foster trust, accountability, and social responsibility in the AI industry, safeguarding against adverse impacts on individuals and society.

5. Ratika is a team leader at a tech company working on an AI project. She is known for her ability to clearly explain complex technical concepts to her team members, ensuring everyone understands their tasks and the project goals. Ratika's excellent communication skills help her team work efficiently and collaboratively. Which soft skill is Ratika demonstrating in her role as a team leader?

Ans. Ratika's exceptional communication skills as a team leader enable her to convey complex technical concepts in a clear, concise, and engaging manner. She adeptly translates technical jargon into understandable language, ensuring alignment, and comprehension among team members with diverse backgrounds and expertise. Ratika fosters open dialogue, active listening, and constructive feedback, creating a supportive environment where ideas are shared, questions are encouraged, and concerns are addressed. Her effective communication enhances collaboration, problem-solving, and decision-making within the team, driving innovation and achieving project objectives. Ratika's ability to articulate vision, goals, and expectations inspires confidence and unity, empowering her team to overcome challenges and excel in their endeavours.

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. What is a key responsibility of a Geographic Information Systems (GIS) Specialist?

a. Developing AI ethics policies ◯ b. Building and maintaining robots ◯

c. Analysing spatial data using AI ◯ d. Writing shell scripts ◯

2. Which skill is crucial for creating and training neural networks in AI?

a. Shell scripting ◯ b. Deep learning ◯

c. Data visualisation ◯ d. Communication skills ◯

3. What role does a Machine Learning Engineer primarily focus on?

a. Designing robotic hardware ◯

b. Creating and training machine learning models ◯

c. Ensuring ethical use of AI ◯

d. Conducting market analysis ◯

4. Which tool is known for data visualisation and is used by AI professionals?

a. TensorFlow ◯ b. Power BI ◯

c. Python ◯ d. C++ ◯

5. Which AI job role involves ensuring responsible development of AI technologies?

a. AI Engineer ◯ b. AI Ethicist ◯

c. Data Scientist ◯ d. Robotics Engineer ◯

6. What is a common application of AI in the healthcare industry?

a. Customer segmentation ◯ b. Fraud detection ◯

c. Medical imaging analysis ◯ d. Travel itinerary planning ◯

7. Which of the following industries uses AI for fraud detection and market analysis?

a. Textile ◯ b. Agriculture ◯

c. Finance ◯ d. Tourism ◯

8. Which platform offers a community experience for data practitioners and provides the latest information on technology, tools, and trends?

a. Kaggle ◯ b. Data Science Central ◯

c. Codecademy ◯ d. Udemy ◯

9. What subject knowledge is essential for an AI-powered Fabric Design Specialist in the textile industry?

a. Geography ◯ b. Chemistry ◯

c. Biology ◯ d. Geology ◯

10. Which python-based package is used for scientific computing and advanced mathematical operations?

a. NumPy ◯ b. SciPy ◯

c. TensorFlow ◯ d. Java ◯

B. Fill in the blanks.

1. AI professionals need to be proficient in using frameworks and libraries such as TensorFlow, SciPy, and _____.

2. In the era of AI, important soft skills include problem-solving abilities, critical thinking, and effective _____.

3. The field of _____ uses AI to develop algorithms for self-driving cars.

4. Defence Contractor creates AI-enabled _____.

5. _____ is a news platform devoted to providing analysis, insight, trends, opinions, and the latest information about big data trends and solutions.

C. State whether the following statement is True or False.

1. AI can only be applied in technical fields and has no impact on creative industries like media and design. _____
2. Data Scientists are responsible for analysing and interpreting complex data to derive insights in AI projects. _____
3. Shell scripting is an important skill for automating tasks and enhancing workflow efficiency in AI development. _____
4. There is no use of AI algorithms in financial sector. _____
5. Communication skills are not important for AI professionals as their work is primarily technical. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. List five common job roles in AI?
2. Why is Python a popular programming language for AI development?
3. What role does a Natural Language Processing (NLP) Specialist play in AI?
4. What is KDnuggets?
5. Mention some essential technical skills for an AI professional.

B. Long answer type questions.

1. Discuss the role of AI in the financial services industry and its impact on operations and decision-making.
2. Explain how AI can be used in the retail sector to enhance customer experiences and optimise business operations.
3. Describe the potential applications of AI in the government and military sectors and their implications.
4. Discuss the significance of base skills for career success in the era of AI.
5. Examine the role of AI in shaping the future of agriculture and its potential impact on food security and sustainability.

C. Competency-based/Application-based questions:

1. Akshat is a Machine Learning Engineer at a tech startup. His current project involves creating and training models for a new AI application that predicts customer preferences and suggests products they might like based on their browsing history. Which specific responsibility aligns with Akshat's job role?
2. Trisha is a Robotics Engineer working at a company that develops AI-powered robots for industrial applications. She is currently designing a robotic arm that can be used in manufacturing to perform precise and repetitive tasks with high efficiency. What is Trisha's main role in her job?
3. Shivani is faced with a complex issue in her AI development project. She breaks down the problem into smaller parts, analyses each one carefully, and then devises a solution that addresses all aspects of the issue. Her systematic approach ensures that the problem is thoroughly understood and effectively resolved. Which soft skill is Shivani demonstrating in her approach to solving the issue?
4. Hemant is working as a Travel Recommendation Engine Developer for a travel agency. His job involves using AI to offer personalised travel recommendations to customers based on their preferences and past travel history. What is Hemant's main responsibility in his job?
5. Vikas is analysing a large dataset for his AI project. He needs to draw meaningful conclusions from the data to inform the development of a new algorithm. Vikas applies logical reasoning and critical analysis to interpret the data accurately, identifying trends and patterns that others might overlook. Which soft skill is Vikas demonstrating in his analysis?

AI In Life

Interdisciplinary

Philosophy also plays a significant role in AI. Find out how and present your findings in class.

AI Deep Thinking

Creativity and Innovativeness

"Whatever happens with the rise of AI, the following skills will always be in demand. Creativity: To be human is to be creative, and while AI offers you useful shortcuts in editing and transcribing...the visual skills of design, illustration and artistry are there to ensure the human touch." Share your views what according to you is the possibility of replacing creativity by AI.

AI Lab

Experiential Learning

1. Use AI to write your resume! Go to https://app.wordtune.com/ and start typing. (You will be asked to login through your Gmail account.) Use AI features to rewrite, shorten or expand items in your resume. You can even adopt a casual or formal tone. Happy writing!
2. Assume you are running a business. Use AI to design a logo for your brand! Go to https://www.logoai.com/logo-maker. Just enter your logo name, your choice of colours and fonts and AI creates your logo for you!

Answers

Exercise (Section A)

A. 1. b 2. b 3. c 4. b 5. c 6. c 7. b 8. b 9. b 10. c

B. 1. social media 2. base 3. Data Scientist
4. Signal processing 5. Computer Vision Engineer

C. 1. True 2. False 3. True 4. False 5. True

Answer the following questions.

Read the article at the given link and answer the questions followed:

https://economictimes.indiatimes.com/jobs/mid-career/ai-could-impact-over-half-of-indian-jobs-linkedin-top-exec/articleshow/108029206.cms?from=mdr

1. Name three soft skills that are in demand for AI jobs in India. Why do you think they are required for a technical job?

2. Identify two important AI technologies desired by any AI professional. Explain why?

3. How is AI useful for recruiters and hirers?

UNIT-3

PYTHON PROGRAMMING

Learning Outcomes

- Features of Python
- Variety of Python Editors
- Modes in Python
- Comments in Python
- Tokens
- Data Types in Python
- Control Statements
- Understanding Libraries
- Introduction to Pandas
- IBM Skill Build: Python for Data Science
- Downloading and Installing Python
- Jupyter Notebook
- Input and Output
- Character Set
- Variables
- Operations on Lists, Tuples and Dictionaries
- CSV Files
- Introduction to NumPy
- Introduction to Scikit-learn

Python is a versatile and dynamic programming language renowned for its simplicity, readability, and extensive range of applications. It was created by Guido van Rossum and first released in 1991. Python got its name from a BBC comedy series – "Monty Python's Flying Circus". Python has rapidly risen to prominence, becoming a staple in various domains such as web development, data science, artificial intelligence, automation, and more. Its clear syntax and easy-to-learn structure make it an ideal choice for beginners, while its powerful libraries and frameworks attract professional developers worldwide. Whether making a web application, analysing complex data sets, or building machine learning models, Python offers a robust and intuitive platform that empowers developers to bring their ideas to life efficiently and effectively.

Features of Python

Python is known for several key features that make it popular among developers. Some of the important features of Python are as follows:

- **Easy to read and write:** Python's syntax is designed to be simple and easy to understand, making it accessible to beginners and experienced programmers. Its code is often referred to as "executable pseudocode."
- **Interpreted language:** Python is an interpreted language, meaning that code is executed line by line, which allows for easier debugging and prototyping.
- **Dynamic typing:** Python is a dynamically typed language. You don't need to declare the type of a variable when you create one. The type of variable is inferred at runtime.
- **Extensive standard library:** Python comes with a large standard library that provides a wide range of modules and packages for tasks such as file I/O, networking, web development, and more, reducing the need to write code from scratch.

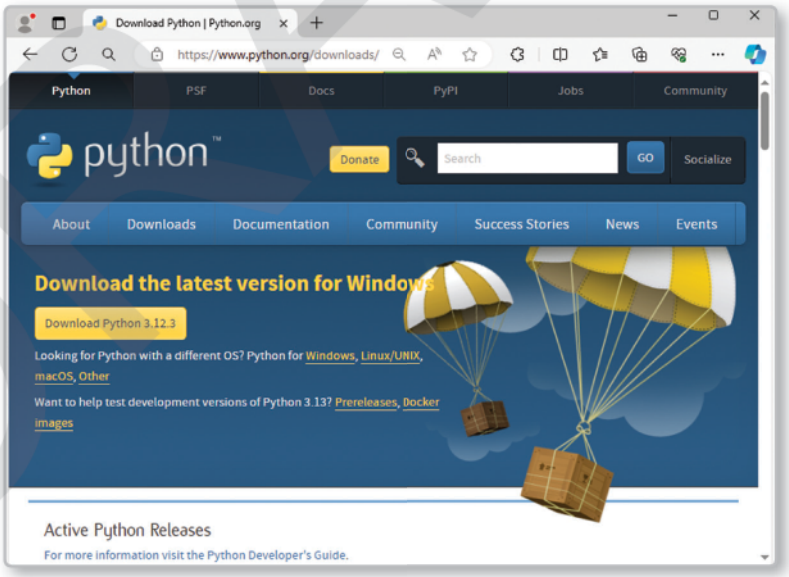
- **Open source:** Python is an open source language. Its source code is freely available and can be modified and redistributed. This fosters collaboration and innovation within the Python community.
- **Object-oriented:** Python supports object-oriented programming (OOP) programming paradigm allowing you to create classes and objects, encapsulate data, and implement inheritance and polymorphism.
- **Cross-platform:** Python code can run on various platforms and operating systems without modification, including Windows, macOS, Linux, and others.
- **Community support:** Python has a vast and active community of developers who contribute to libraries, frameworks, and resources. This community support makes it easy to find solutions to problems and learn from others.
- **Easy integration:** Python effortlessly integrates with other programming languages and technologies, facilitating interoperability and compatibility with existing systems. It supports integration with C/C++, Java, .NET, and other languages that allows developers to use their preferred tools and technologies within Python-based projects.
- **Strong industry adoption:** Python is widely adopted across various industries, including technology, finance, healthcare, education, and more. Its robustness, scalability, and extensive library support make it an attractive choice for businesses seeking efficient and cost-effective solutions to their software development needs.
- **High-level language:** Python is a high-level language, meaning it abstracts much of the complex details of the computer's hardware. This allows developers to focus more on solving problems rather than worrying about the underlying system architecture.
- **Comprehensive character processing:** Python can handle all ASCII and Unicode characters, making it suitable for international applications that require the processing of a variety of character sets and symbols.

Downloading and Installing Python

The steps to download and install Python are as follows:

Step 1: Visit the official Python website at https://www.python.org/downloads/.

The download button of the latest version of Python display on the website.



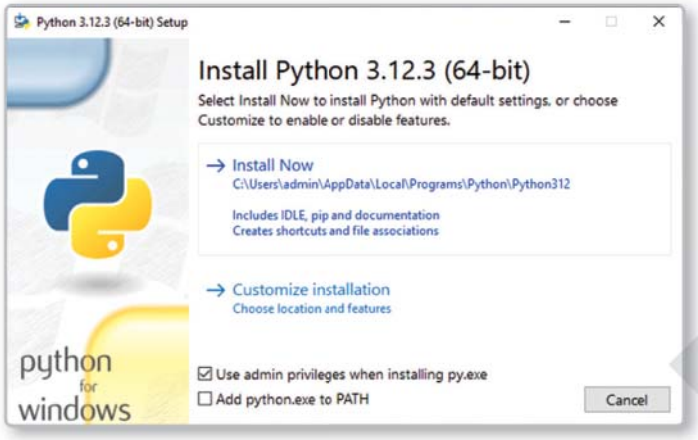
Step 2: Click on the **Download** button. The download should start automatically. Wait for the download to complete.

Step 3: After downloading the installer, locate the file and run it by double-clicking on it.

The process of installing the Python begins.

Step 4: Click on the **Add python.exe to PATH** check box to set Python to the Environment variable of the operating system.

Step 5: Click on the **Install Now** button.



Step 6: Follow the instruction given on the Installation Wizard that will guide you through the installation process.

After completing the installation process, Python will be installed on your system.

Brainy Fact

- Python is widely used by numerous major companies, including Google, Facebook, and Amazon.

Variety of Python Editors

Python offers a wide range of editors and integrated development environments (IDEs) that cater to different needs and preferences. Some of the popular ones are described as follows:

- **Python IDLE:** IDLE (Integrated Development and Learning Environment) is Python's default IDE that comes bundled with the standard distribution of Python.
- **PyCharm:** PyCharm is a professional IDE developed by JetBrains, available in both free (Community) and paid (Professional) versions.
- **Anaconda:** Anaconda is a distribution of Python and R for scientific computing and data science. It includes a package manager (conda) and many pre-installed libraries.
- **Spyder:** Spyder (Scientific Python Development Environment) is an open-source IDE that comes with Anaconda.
- **Jupyter Notebook:** Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualisations, and narrative text.

Let us learn about Jupyter Notebook in detail.

Jupyter Notebook

Jupyter Notebook is a powerful and versatile tool that revolutionises the way programmers and data scientists work with code and data. By providing an interactive environment where code execution and rich media can coexist. Jupyter Notebook facilitates a more dynamic and intuitive approach to programming. Users can write and run code in segments, view the results immediately, and combine these outputs with formatted text and visualisations. It is widely used in data science, scientific computing, and education.

Features of Jupyter Notebook are as follows

- **Interactive Computing—Live Code Execution:** Write and execute code in real-time, and see the results immediately within the same document.

- **Data Visualization—Integrated Plots:** Easily create and display visualisations using libraries like Matplotlib, Seaborn, Plotly, and Bokeh directly within the notebook.
- **Collaboration and Sharing—Export Formats:** Export notebooks to various formats, including HTML, PDF, Markdown, and slides.

You can install Jupyter Notebook using pip. For installing Jupyter Notebook, you need to open your terminal or command prompt and run the following command:

```
pip install notebook
```

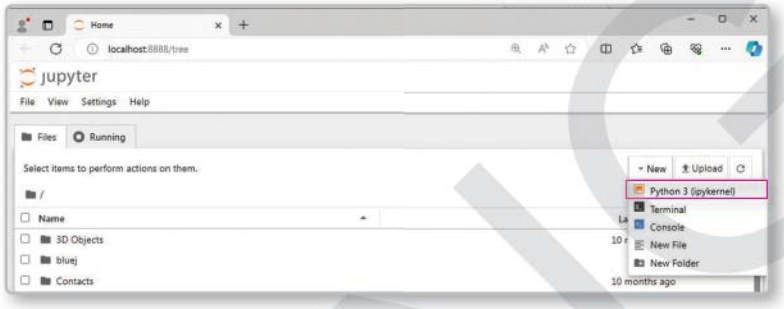
The steps to open and work with Jupyter Notebook are as follows:

Step 1: Open the Jupyter Notebook by running the following command in your terminal or command prompt:

```
jupyter notebook
```

This will start the Jupyter server and open the Jupyter dashboard in your default web browser.

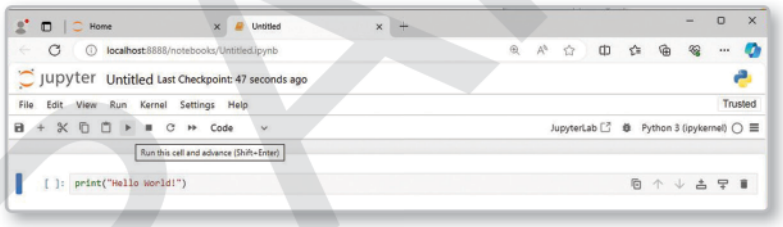
Step 2: Click the **New** button and select the **Python 3** option in the Jupyter dashboard to create a new notebook.



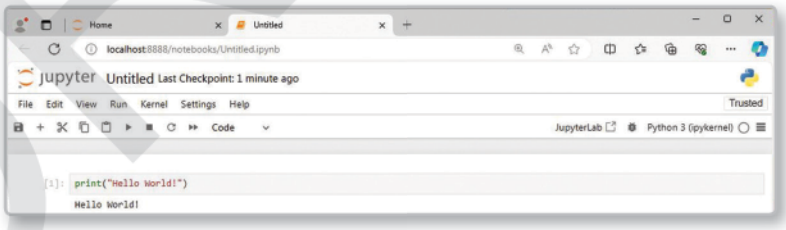
The new notebook opens with the default name Untitled. However, you can change the name by clicking over it. Jupyter Notebook consists of cells. The two main types are Code cells and Markdown cells.

Step 3: Write the code in the Code cell.

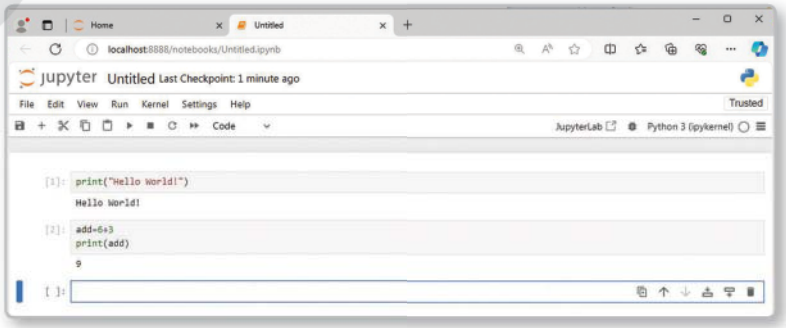
Step 4: Click on the **Run** button in the toolbar or press the **Shift + Enter** key to run a code.



The output of the code is displayed.



We can work on multiple programs in the same window.



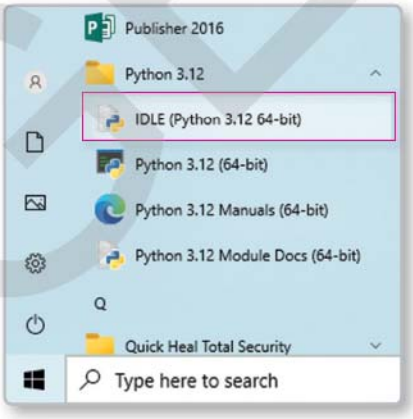
You can also use an online Jupyter Notebook through Google Colab (This requires login through your Gmail account).

Modes in Python

In Python, particularly when using Python IDLE, there are two main modes in which you can work. These modes are Interactive mode and Script mode.

Interactive Mode

Interactive mode allows you to enter Python commands directly into the command line shell and immediately see the results. In Interactive mode, we type one command at a time. Interactive mode is useful for quick experimentation, testing snippets of code, or exploring the behaviour of Python functions and modules. The Interactive mode does not save the commands entered by the user in the form of a program.



Follow these steps to start Python IDLE and work in Interactive mode:

Step 1: Click on the **Start** button.

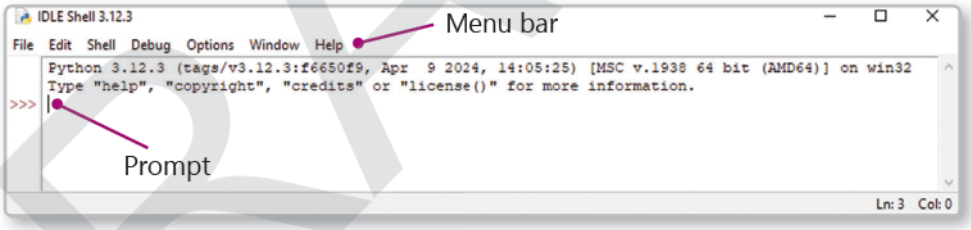
Step 2: Scroll down the list of programs and click on the **Python 3.12** folder.

Step 3: Click on the **IDLE (Python 3.12 64-bit)** app.

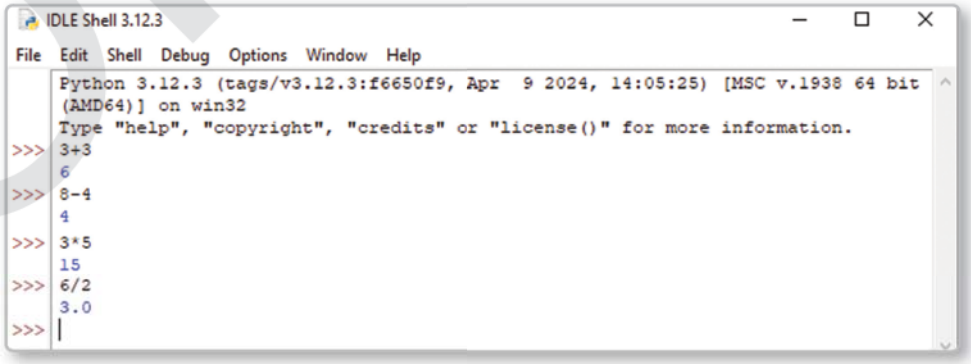
The main IDLE window appears. It is a code editor which helps you to write and execute Python programs in the editor itself.

There are two components of Python IDLE window:

- **Menu Bar:** The Menu Bar of Python IDLE window is similar to the Menu Bar of other programs. It has various menus such as File, Edit, Shell, Debug, Options, Window and Help.
- **Prompt:** You will see symbol (>>>) in the window. This is known as the Prompt. The Prompt allows the user to enter commands directly into Python and get an output instantly by pressing the **Enter** key.



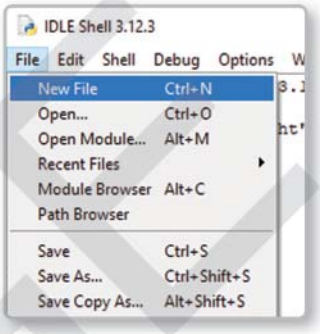
Type the commands at the prompt and you will immediately get the results. Simply, type **3 + 3** and press the Enter key. The result **6** appears on the screen. Try more values with different operations and see result.



Script Mode

Script mode refers to writing and running Python programs as standalone scripts, rather than executing code interactively. Script mode involves writing Python code in a text file with a .py extension and then executing that file using the Python interpreter. You create Python scripts using a text editor or an Integrated Development Environment (IDE) and then run them from the command line or terminal. Script mode is suitable for developing larger programs, modules, or scripts that perform specific tasks or functions.

Create a New File



Follow these steps to create a new file:

Step 1: Open the Python IDLE window.

Step 2: Click on the **File** menu. A drop-down menu appears.

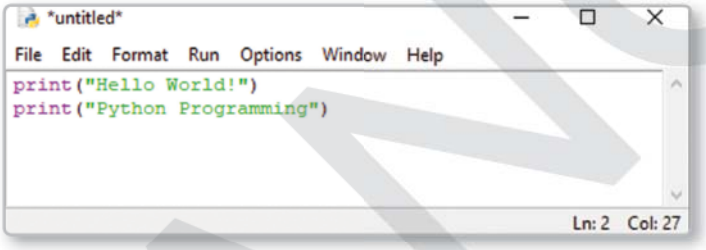
Step 3: Click on the **New File** option.

A new editor window with a blank file will appear on the screen.

Create and Save a Program

The steps to write and save a program are as follows:

Step 1: Type the program code in Python window.



Step 2: Click on the **File** menu.

Step 3: Click on the **Save** option. The **Save As** dialog box appears.

Step 4: Navigate the location where you want to save the file.

Step 5: Enter the name of the file in the **File name** text box.

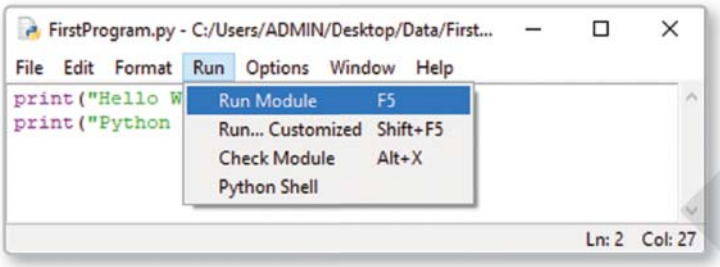
Step 6: Click on the **Save** button.

Run a Program

The file will be saved by the given name at the specified location. Follow these steps to run a Python program:

Step 1: Click on the **Run** menu in the **Menu** bar. Alternately, press F5 to run the program.

Step 2: Click on the **Run Module** option.



The program will execute and the output will be shown in the **Shell** window.

Input and Output

Python provides two commonly used functions `input()` and `print()` for input and output.

- **The input() Function:** The `input()` function takes the user's input while a program executes. When called, it waits for the user to type something on the keyboard and press the Enter key. The function then returns the user input as a string. The syntax of the `input()` function is as follows:

  ```
  input([<prompt>])
  ```

  Where, prompt is the string or message we wish to display on the screen.

  Example:

  ```
  name = input("Enter your name: ")
  ```

- **The print() Function:** The `print()` function prints or sends the output to the standard output device, which is usually a monitor. This function auto converts the items to strings, i.e., if you try printing a numeric value, the `print()` function will automatically convert it into equivalent string and print it. The `print()` function also introduces spaces between items automatically with the help of the comma operator.

  The syntax of the `print()` function is as follows:

  ```
  print (<expression 1> , <expression 2>...)
  ```

  Example:

  ```
  print ("Python is easy")
  ```

Program 1: To demonstrate the use of the input( ) and print( ) functions

```
name =input("Enter your name: ")
age =input("Enter your age: ")
print(name,"your age is", age)
```

Output:

```
Enter your name: Yash
Enter your age: 25
Yash your age is 25
```

Comments in Python

In Python, comments are used to annotate code with explanations, documentation, or notes. Comments are ignored by the Python interpreter during execution and are solely for human readers. Python supports two types of comments:

- **Single line comments:** Single line comments start with the hash character # and continue until the end of the line. Anything following the # on that line is considered a comment and is ignored by the interpreter.

Program 2: To demonstrate the use of single line comment

```
# This is a single line comment
print("Hello, world!")  # This is another single line comment
```

Output:

```
Hello, world!
```

- **Multiline comments:** Python does not have a built-in syntax for multiline comments like some other languages. However, multiline strings enclosed within triple quotes (''' or """) are often used for this purpose. Although these are technically strings and not comments, they serve the purpose of commenting out multiple lines of code.

Program 3: To demonstrate the use of multiline comments

```
'''
This program demonstrates basic arithmetic operations without using functions.
It performs addition, subtraction, multiplication, and division of two numbers.
'''
# Uses of Arithmetic Operators

# Addition
a = 10
b = 5
result_add = a + b
print("Addition:", result_add)

# Subtraction
result_sub = a - b
print("Subtraction:", result_sub)

# Multiplication
result_mul = a * b
print("Multiplication:", result_mul)

# Division
result_div = a / b
print("Division:", result_div)
```

Output:

```
Addition: 15
Subtraction: 5
Multiplication: 50
Division: 2.0
```

Character Set

In Python, a character set refers to a collection of characters, typically defined by a specific encoding scheme, such as ASCII (American Standard Code for Information Interchange) or Unicode. Each character in a character set is represented by a unique code point.

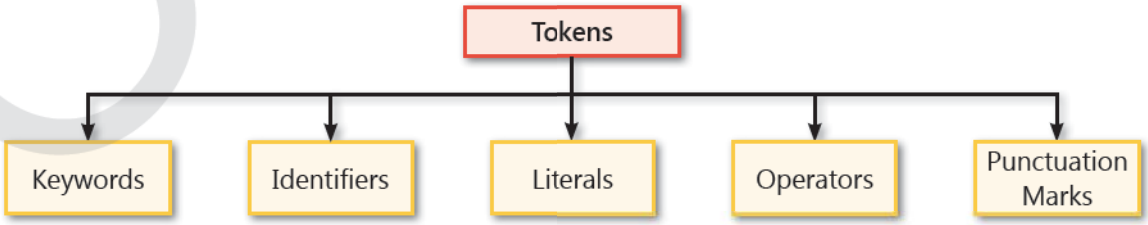
A programming language's character set refers to the permissible characters recognised by that language. When discussing the Python programming language specifically, its character set encompasses all valid characters allowed for scripting. These include:

- **Letters:** Both uppercase (A-Z) and lowercase (a-z) letters.
- **Digits:** All numerical digits from 0 to 9.
- **Special Symbols:** Python accommodates various special symbols such as single quotes ('...'), double ("..."), semicolons (;), colons (:), exclamation marks(!), tildes (~), at symbols(@), hash signs (#), dollar signs ($), percentage signs (%), caret sign (^), backticks (`), ampersands (&), asterisks (*), parentheses (), underscores(_), plus signs(+), hyphens(-), equal signs(=), curly braces {}, square brackets [], and backslashes(\).
- **Whitespace characters:** Python recognises whitespace characters including tab spaces, blank spaces, newline characters, and carriage return characters.
- **Others:** Python supports the entirety of ASCII and Unicode characters, comprising the complete Python character set.

Tokens

A token is the smallest unit of a program that has a meaning. Tokens in Python are fundamental elements of the language's syntax. They represent the building blocks of Python code, consisting of keywords, identifiers, literals, operators, and punctuations. Understanding tokens is essential for comprehending Python's lexical structure (basic syntax and set of rules defining how Python programs are written) and parsing rules (guidelines used by a parser to analyse the structure of code).

Python code is first divided into tokens during the lexical analysis phase of the interpretation process. These tokens are then used by the parser to construct the abstract syntax tree (AST a tree representation of Python code that is used for code analysis and manipulation), which is further processed to execute the code.



Keywords

Keywords are predefined and reserved words in Python, which have special meanings and purposes. They cannot be used as identifiers or variable names. These keywords are part of Python's syntax and are used to define control flow, declare functions and classes, handle exceptions, etc.

Following table shows the keywords used in Python:

and	as	assert	async	await
break	class	continue	def	del
elif	else	except	False	finally
for	from	global	if	import
in	is	lambda	None	nonlocal
not	or	pass	raise	return
True	try	while	with	yield

Identifiers

Identifiers are the names used to identify variables, functions, classes, modules, and other objects in Python. They act as labels for these elements, allowing you to refer to them in your code. Identifiers follow certain rules and conventions:

- Must begin with a letter (a-z, A-Z) or an underscore (_).
- Subsequent characters can be letters, digits (0-9), or underscores.
- Case-sensitive (`myVar` is different from `myvar`).
- Cannot be a Python keyword (reserved words).
- No special characters such as !, @, #, $, %, etc., are allowed within identifiers.
- Blank spaces within an identifier are disallowed.

For example:

```
my_variable
calculate_sum
MyClass
```

Some conventions that can be used while specifying identifiers for more clarity are as follows:

- Use descriptive names that convey the purpose of the variable, function, etc.
- Follow the **snake_case** convention for variable names and function names. This means using all lowercase letters with words separated by underscores. For example: my_variable, calculate_sum.
- Follow the **PascalCase** convention for class names. This means capitalising the first letter of each word, without underscores. For example: MyClass, BankAccount.
- Use all uppercase letters with underscores to denote constants. For example: MAX_SIZE, PI.

Literals

Literals in Python represent fixed values that are used directly in the code. They can be of several types:

- **Numeric literals:** Represented as integers, floating-point numbers, or complex numbers. Examples: 42, 3.14, 1 + 2j.
- **String literals:** Represented as sequences of characters enclosed in single, double, or triple quotes. Examples: 'hello', "world", '''multiline string'''.
- **Boolean literals:** Represented as True or False.
- **None literal:** Represented as None that indicates the absence of a value.

Operators

In Python, operators can be defined as special symbols which perform arithmetic and logical computation. The values which the operators use to get the output are called **operands**.

Arithmetic Operators

Arithmetic operators are used to perform mathematical operations such as addition, subtraction, multiplication, division, modulus, exponentiation, and floor division.

Operator	Description
+	Adds two operands or unary plus
-	Subtracts right operand from left operand or unary minus
*	Multiplies two operands
/	Divides left operand by right operand
%	Returns remainder of division
//	Performs floor division—division that results into whole number adjusted to the left in the number line
**	Exponentiation—left operand raised to the power of right operand

Program 4: To demonstrate the use of arithmetic operators

```
#The input() function returns a string
#So we use the int() function to convert the strings to integers

num1 = int(input("Enter first number: "))
num2 = int(input("Enter second number: "))

# Perform arithmetic operations
addition = num1 + num2
subtraction = num1 - num2
multiplication = num1 * num2
division = num1 / num2
modulus = num1 % num2
exponentiation = num1 ** num2
floor_division = num1 // num2

# Display the results
print(num1, "+", num2, "=", addition)
print(num1, "-", num2, "=", subtraction)
print(num1, "*", num2, "=", multiplication)
print(num1, "/", num2, "=", division)
print(num1, "%", num2, "=", modulus)
print(num1, "**", num2, "=", exponentiation)
print(num1, "//", num2, "=", floor_division)
```

Output:

```
Enter the first number: 16
Enter the second number: 3
16 + 3 = 19
```

```
16 - 3 = 13
16 * 3 = 48
16 / 3 = 5.333333333333333
16 % 3 = 1
16 ** 3 = 4096
16 // 3 = 5
```

Comparison Operators

Comparison operators are used to compare two values and return a boolean result (True or False). They return True if the comparison is true, otherwise False. They are commonly used in conditional statements and loops.

Operator	Description
==	Checks if two operands are equal
!=	Checks if two operands are not equal
>	Checks if left operand is greater than right operand
<	Checks if left operand is less than right operand
>=	Checks if left operand is greater than or equal to right operand
<=	Checks if left operand is less than or equal to right operand

Program 5: To demonstrate the use of comparison operators

```
# Uses of Comparison Operators

# Equal to
print("Equal to:", 10 == 5)

# Not equal to
print("Not equal to:", 10 != 5)

# Greater than
print("Greater than:", 10 > 5)

# Less than
print("Less than:", 10 < 5)

# Greater than or equal to
print("Greater than or equal to:", 10 >= 5)

# Less than or equal to
print("Less than or equal to:", 10 <= 5)
```

Output:

```
Equal to: False
Not equal to: True
Greater than: True
Less than: False
Greater than or equal to: True
Less than or equal to: False
```

Assignment Operators

Assignment operators are used to assign values to variables. They combine the assignment operation with another operation such as addition, subtraction, multiplication, etc.

Operator	Description
=	Assigns value of right operand to left operand
+=	Adds right operand to left operand and assigns the result to left operand
-=	Subtracts right operand from left operand and assigns the result to left operand
*=	Multiplies left operand by right operand and assigns the result to left operand
/=	Divides left operand by right operand and assigns the result to left operand
%=	Computes modulus of left operand with right operand and assigns the result to left operand
//=	Performs floor division on left operand by right operand and assigns the result to left operand
**=	Raises left operand to the power of right operand and assigns the result to left operand

Program 6: To demonstrate the use of assignment operators

```
# Uses of Assignment Operators

# Assignment
x = 15
print("Assignment:", x)

# Addition assignment
x += 3
print("Addition Assignment:", x)

# Subtraction assignment
x -= 2
print("Subtraction Assignment:", x)

# Multiplication assignment
x *= 4
print("Multiplication Assignment:", x)

# Division assignment
x /= 2
print("Division Assignment:", x)

# Modulus assignment
x %= 3
print("Modulus Assignment:", x)
```


```
# Floor division assignment
x //= 2
print("Floor Division Assignment:", x)

# Exponentiation assignment
x **= 3
print("Exponentiation Assignment:", x)
```

Output:

```
Assignment: 15
Addition Assignment: 18
Subtraction Assignment: 16
Multiplication Assignment: 64
Division Assignment: 32.0
Modulus Assignment: 2.0
Floor Division Assignment: 1.0
Exponentiation Assignment: 1.0
```

Logical Operators

Logical operators are used to combine conditional statements. They return True or False depending on the conditions. Logical operators include AND, OR, and NOT.

Operator	Description
and	Returns True if both operands are true
or	Returns True if at least one operand is true
not	Returns True if operand is false

Program 7: To demonstrate the use of logical operators

```
# Uses of Logical Operators

x = 10
y = 5
z = 20

# Logical AND: True if both conditions are True
logical_and_result = (x > y) and (y < z)
print("Logical AND Result:", logical_and_result)

# Logical OR: True if at least one condition is True
logical_or_result = (x < y) or (y < z)
print("Logical OR Result:", logical_or_result)
```

```
# Logical NOT: Inverts the result, True becomes False and False becomes True
logical_not_result = not (x == y)
print("Logical NOT Result:", logical_not_result)
```

Output:

```
Logical AND Result: True
Logical OR Result: True
Logical NOT Result: True
```

Ai Task

Write Python scripts for the following:

1. Accept a number and display its square and cube.

2. Accept length and breadth of a rectangle. Display its area and perimeter.

3. Accept principal amount, rate of interest and time period. Calculate and display simple interest.

4. Predict the output:

```
print((7 + 3) - (7 + 3))
print(100 + 15 * 2)
```

Bitwise Operators

Bitwise operators perform bit-level operations on operands. They are used to manipulate individual bits of integers.

Operator	Name	Description
&	Bitwise AND	Returns bit 1, if both bits are 1; otherwise returns bit 0.
|	Bitwise OR	Returns bit 1, if any of the bits is 1; otherwise returns bit 0.
^	Bitwise XOR	Returns bit 1, if any of the bits is 1 but not both; otherwise returns bit 0.
~	Bitwise NOT (complement)	Inverts individual bits.
<<	Bitwise left shift	It moves all bits in a binary number to the left by a certain number of positions.
>>	Bitwise right shift	It moves all bits in a binary number to the right by a certain number of positions.

Program 8: To demonstrate the use of bitwise operators

```
# Uses of Bitwise Operators

# Bitwise AND
a = 10  # Binary: 1010
b = 6   # Binary: 0110
result_and = a & b
print("Bitwise AND:", result_and)

# Bitwise OR
result_or = a | b
print("Bitwise OR:", result_or)

# Bitwise XOR
result_xor = a ^ b
print("Bitwise XOR:", result_xor)

# Bitwise NOT
result_not = ~a
print("Bitwise NOT:", result_not)

# Bitwise Left Shift
result_left_shift = a << 2
print("Bitwise Left Shift:", result_left_shift)

# Bitwise Right Shift
result_right_shift = a >> 2
print("Bitwise Right Shift:", result_right_shift)
```

Output:

```
Bitwise AND: 2
Bitwise OR: 14
```

```
Bitwise XOR: 12
Bitwise NOT: -11
Bitwise Left Shift: 40
Bitwise Right Shift: 2
```

Ai Reboot

Answer the following questions:

a. Name any two keywords in Python.

b. What is a token?

c. Which operator is used to represent exponent in Python?

d. Which operator returns the quotient after division as an integer?

Membership Operators

Membership operators are used to test whether a value is present in a sequence (such as a string, list, tuple, etc.) or not.

Operator	Description
in	Returns True if a specific value is present in a sequence
not in	Returns True if a specific value is not present in a sequence

Program 9: To demonstrate the use of membership operators

```
# Uses of Membership Operators

# in operator
my_list = [1, 2, 3, 4, 5]
print("Is 3 in the list?", 3 in my_list)

# not in operator
print("Is 6 not in the list?", 6 not in my_list)
```

Output:

```
Is 3 in the list? True
Is 6 not in the list? True
```

Identity Operators

Identity operators are used to compare the memory locations of two objects rather than their values.

Operator	Description
is	Returns True if both variables are the same object in the memory.
is not	Returns True if both variables are not the same object in the memory.

Program 10: To demonstrate the use of identity operators

```
# Uses of Identity Operators

# is
x = [1, 2, 3]
y = [1, 2, 3]
print("Is x the same object as y?", x is y)

# is not
print("Is x not the same object as y?", x is not y)
```

Output:

```
Is x the same object as y? False
Is x not the same object as y? True
```

Operator Precedence

Precedence of operators determines the order in which the operators are executed. The operator precedence in Python is listed in the following table. The highest precedence is at the top.

Operator	Name
()	Parentheses
**	Exponent
*, /, %, //	Multiplication, Division, Modulus, Floor Division
+, –	Addition, Subtraction
==, !=, >, <, >=, <=	Comparison
=, +=, -=, *=, /=, %=, **=, //=	Assignment
and, or, not	Logical

Punctuation Marks

Punctuation marks in Python are special characters used for various purposes such as accessing attributes, separating elements, specifying syntax, and more. Some common punctuation marks in Python are as follows:

- **Period (.):** Used for accessing attributes and methods of objects.
- **Comma (,):** Separates elements in tuples, lists, function arguments, and dictionary key-value pairs.
- **Colon (:):** Used in dictionaries to separate keys from values, in slices, and in control structures like loops and conditionals. To define the beginning of an indented code block.
- **Semicolon (;):** Can be used to separate statements on the same line.
- **Question Mark (?):** In some contexts, such as libraries like NumPy or IPython, it's used for querying documentation or help.
- **Exclamation Mark (!):** Often used in Python shells like IPython or Jupyter notebooks to execute shell commands.
- **Parentheses (()) and Brackets ([]):** Used for grouping expressions, function calls, and indexing/slicing sequences.
- **Quotation Marks (' and "):** Used to denote string literals.
- **Curly Braces ({}):** Used to denote sets and dictionaries, and in formatting strings.
- **Asterisk (*):** Used for multiplication, exponentiation, and in unpacking iterables.
- **Forward Slash (/):** Used for division.
- **Backslash (\):** Used for line continuation or as an escape character in strings.

- **Percentage Sign (%):** Used in string formatting (old-style formatting).
- **Underscore (_):** Used as a variable name, or in naming conventions, such as snake_case.
- **Hash (#):** Used to create comments in Python code.
- **At Sign (@):** Used in decorators.
- **Dollar Sign ($):** Used in some Python libraries like pandas for referencing columns in data frames.
- **Tilde (~):** Used as the complement operator in bitwise operations.
- **Double Underscore (__):** Used in special methods (dunder/magic methods) and sometimes for name mangling in classes.
- **Angle Brackets (< and >):** Used for comparison operations.
- **Vertical Bar or Pipe (|):** Used in bitwise OR operations.
- **Ampersand (&):** Used in bitwise AND operations.
- **Caret (^):** Used in bitwise XOR operations.
- **Double Period (..):** Used in some libraries and frameworks for specifying ranges or intervals.
- **Ellipsis (...):** Used as a placeholder in some contexts, such as in NumPy arrays or as part of type hints.
- **Minus-Equivalent (-=):** Used in augmented assignment statements.
- **Double Greater Than (>>) and Double Less Than (<<):** Used in bit shifting operations.

Variables

In Python, variables are used to store and manipulate data values. Variables in Python do not have fixed locations. The location they refer to changes every time their values change.

Variable Assignment

Variables in Python are declared simply by assigning a value to them using the assignment operator (=).

Python syntax for assigning a value to a variable can be described as follows:

```
<Variable_name> = <value>
```

For example:

```
x = 10
name = "Yash"
my_list = [1, 2, 3, 4, 5]
```

Unlike some other programming languages, Python does not require explicit declaration of variables before they are used. When a variable is assigned a value, Python automatically determines its data type. Python variables can hold values of various data types, including integers, floats, strings, booleans, lists, tuples, dictionaries, and more.

For example:

```
# Integer variable
age = 30

# Float variable
pi = 3.14

# String variable
message = "Hello, World!"

# Boolean variable
is_student = True
```


```
# List variable
numbers = [1, 2, 3, 4, 5]
# Tuple variable
point = (10, 20)

# Dictionary variable
person = {'name': 'Yash', 'age': 30}
```

Multiple Assignment

Multiple assignment of variables, also known parallel assignment, is a powerful feature in many programming languages that allows you to assign multiple variables at once. This concept can significantly enhance code readability, efficiency, and flexibility.

For example:

```
x, y, z = 1, 2, 3
```

In given code, the values 1, 2, and 3 are assigned to variables x, y, and z, respectively. This simultaneous assignment saves lines of code and improves readability compared to assigning each variable separately.

Multiple assignment is not limited to simple variables. It works with any iterable data structure, including tuples, lists, and even custom objects that support iteration and unpacking.

For example:

```
point = (5, 10)
x, y = point
```

In given code, the values (5, 10) are unpacked from the point tuple and assigned to variables x and y respectively. This syntax is concise and expressive, making code more elegant and maintainable.

For example:

```
[a, b, c] = [1, 2, 3]
```

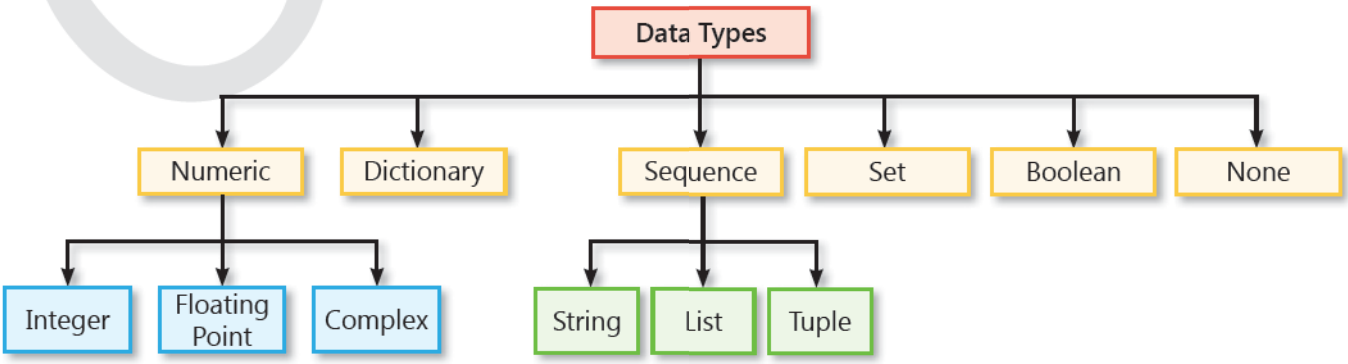
In this example, a, b, and c are assigned the values 1, 2, and 3 respectively from the list [1, 2, 3].

Data Types in Python

Data types in Python specify the type of values that a variable can hold. Python is dynamically typed, which means there is no need to declare the data type of a variable explicitly. Python automatically assigns the appropriate data type to a variable based on the value assigned to it.

This flexibility makes Python a versatile language for various programming tasks. Python allows different operations to be performed on different data types.

The following tree structure shows the various data types used in Python:



Numeric Data Type

Numeric data types are used to store numeric values. In Python, there are three main numeric data types: integer, floating point, and complex.

- **Integer (int):** Integers are whole numbers, positive or negative, without any decimal point. They can be of unlimited size, subject only to the memory available.

  For example:

  ```
  x = 10
  y = -20
  ```

- **Floating Point (float):** Floating-point numbers (floats) represent real numbers and are written with a decimal point dividing the integer and fractional parts.

  For example:

  ```
  pi = 3.14
  temperature = 98.6
  ```

- **Complex (complex):** Complex numbers are written with a "j" as the imaginary part. A complex number is of the form x + yj and comprises a pair of floating point numbers. The first part (x) is called the real part, and the second one (y) is called the imaginary part.

  For example:

  ```
  z = 5 + 3j
  ```

Dictionary Data Type

Dictionaries are unordered collections of key-value pairs enclosed in curly braces {}. Each key in a dictionary must be unique and immutable, and it is used to access its corresponding value. Dictionaries are commonly used because they provide fast access to values based on their keys.

For example:

```
person = {'name': 'John', 'age': 30, 'city': 'New York'}
```

Sequence Data Type

In Python, a sequence data type refers to an ordered collection of elements, where each element is indexed by a non-negative integer. There are mainly three sequence data types in Python, including lists, tuples, and strings.

- **Lists:** Lists are mutable sequences, meaning their elements can be changed after creation. They are defined using square brackets [ ] and can contain elements of different data types.

  For example:

  ```
  my_list = [1, 2, 3, 'a', 'b', 'c']
  ```

- **Tuples:** Tuples are immutable sequences, meaning their elements cannot be changed after creation. They are defined using parentheses ( ) and can contain elements of different data types.

  For example:

  ```
  my_tuple = (1, 2, 3, 'a', 'b', 'c')
  ```

- **Strings:** Strings are immutable sequences of characters. They are defined using either single quotes (' ') or double quotes (" ").

  For example:

  ```
  my_string = 'Hello, World!'
  my_string1 = "Hello, World!"
  ```

  In Python, you can create multiline strings using triple quotes. Triple quotes can be either single quotes (''') or double quotes (""""). Multiline strings are often used for docstrings (documentation strings), multiline comments, or for storing large blocks of text.

For example:

```
multiline_string = """
This is a multiline string.
It can span across multiple lines.
Triple quotes allow you to include line breaks without using escape characters.
"""
```

Set Data Type

In Python, a set is an unordered collection of unique elements. Sets are mutable, meaning you can add or remove elements from them after they are created. You can create a set in Python by enclosing a comma-separated list of elements within curly braces {}. The key characteristics of sets in Python is that they do not allow duplicate elements.

For example:

```
# Creating a set using curly braces
my_set = {1, 2, 3, 4, 5}
```

Boolean Data Type

The Boolean data type in Python represents one of two values: `True` or `False`. Booleans are commonly used for logical operations, conditional expressions, and control flow in programming. In Python, the integer value 0 is considered false, while any non-zero integer value is considered true. This means that when evaluated in a boolean context, 0 evaluates to False, and any other non-zero integer value evaluates to True.

For example:

```
is_valid = True
is_empty = False
```

None Data Type

None is a special constant in Python that represents the absence of a value or represents null value. It is often used to indicate that a variable has not been assigned a value.

For example:

```
x = None
```

Type Casting

Type casting, also known as type conversion, in Python refers to the process of converting a variable from one data type to another. Python has two types of type casting: implicit and explicit. In case of implicit type casting, Python automatically converts one data type to another data type. On the other hand, we need to convert one data type to another data type explicitly by using the Python functions. Python provides several built-in functions for type casting. Some of them are as follows:

- **`int()`:** Converts a value to an integer.
- **`float()`:** Converts a value to a float.
- **`str()`:** Converts a value to a string.
- **`bool()`:** Converts a value to a boolean.
- **`list()`:** Converts a value to a list.
- **`tuple()`:** Converts a value to a tuple.
- **`set()`:** Converts a value to a set.
- **`dict()`:** Converts a sequence of key-value pairs to a dictionary.

Program 11: To demonstrate the concept of type casting

```
# Demonstrating type casting in Python
# Initial values of different types
```

```
int_value = 10
float_value = 15.5
str_value = "25"
bool_value = True

# Type casting
# Convert integer to float
int_to_float = float(int_value)
# Convert float to integer
float_to_int = int(float_value)
# Convert string to integer
str_to_int = int(str_value)
# Convert boolean to integer
bool_to_int = int(bool_value)

# Print the casted values and their new types
print("Casted values and their new types:")
print("int_to_float:", int_to_float, type(int_to_float))
print("float_to_int:", float_to_int, type(float_to_int))
print("str_to_int:", str_to_int, type(str_to_int))
print("bool_to_int:", bool_to_int, type(bool_to_int))
```

Output:

```
Casted values and their new types:
int_to_float: 10.0 <class 'float'>
float_to_int: 15 <class 'int'>
str_to_int: 25 <class 'int'>
bool_to_int: 1 <class 'int'>
```

The `type()` function is used to display the data type of a variable or object.

Operations on Lists

We can perform different types of operations on a list, like accessing elements, adding elements, and removing elements. Before performing any operation, we need to first create a list.

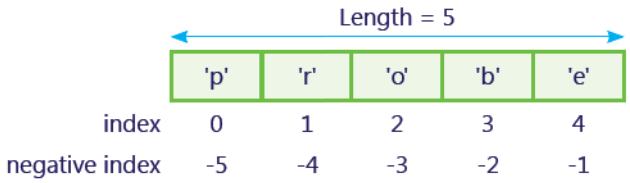
Creating a List

You can create a list by placing all the elements inside square brackets [], separated by commas.

```
L=[]          # creates an empty list
L=[1, 2, 3, 4, 5]         # creates a list with numeric values
fruits = ["kiwi", "pineapple", "cherry", "orange"] # creates a list with string values
mix=["India", 12, 34.5]   # creates a list with mixed data types
nested_list[[1, 2, 3], 12, 5, 6]      # creates a nested list
A nested list is a type of list that contains another list as its element.
```

Indexing in Lists

Lists are indexed, which means each element has a unique position. The index starts at 0 for the first element, 1 for the second, and so on. Negative indexing starts from the right to left with -1 being the last element.



In order to access an element in a list we need to use index operator [].

```
# Accessing elements using positive indexing
print(fruits[0])  # Output: kiwi
print(fruits[1])  # Output: pineapple

# Accessing elements using negative indexing
print(fruits[-1])  # Output: orange
print(fruits[-2])  # Output: cherry
```

Lists are mutable, so you can change their elements. Just specify the index value with the new element to be added.

```
fruits[1] = "blueberry"
print(fruits)  # Output: fruits = ["kiwi", "blueberry", "cherry", "orange"]
```

Adding Elements

You can add elements to a list by using the `append()`, `insert()`, and `extend()` methods.

Using append() Function

You can only add one element at a time with `append()`. If you want to add multiple elements using `append()`, you need to use loops. Tuples can also be added to a list using `append()` because tuples cannot be changed after they're created. Unlike sets, you can add lists to an existing list using the `append()` function.

```
# Adding elements using append() function
fruits.append("strawberry")
print(fruits)
```

Output:

```
["kiwi", "blueberry", "cherry", "orange", "strawberry"]
```

Using insert() Function

The `append()` function adds an element only to the end of a list. To add an element at a specific position, you need to use the `insert()` function. Unlike `append()`, which takes only one argument (the value to add), `insert()` requires two arguments: the position where you want to insert the element and the value itself.

```
fruits=["kiwi", "blueberry", "cherry", "orange", "strawberry"]
fruits.insert(2, "banana")
print(fruits)
```

Output:

```
['kiwi', 'blueberry', 'banana', 'cherry', 'orange', 'strawberry']
```

Using extend() Function

The extend() method is used to add multiple elements to the end of a list at the same time.

```
# Adding elements using extend() function
fruits1=['apple', 'mango']
fruits.extend(fruits1)
print(fruits)
```

Output:

```
['kiwi', 'blueberry', 'banana', 'cherry', 'orange', 'strawberry', 'apple', 'mango']
```

Removing List Elements

You can remove elements from the list by using 'remove' and 'pop' functions.

Using remove() Function

The remove function deletes the first occurrence of the specified value from the list.

```
# Removing element using remove() function
fruits.remove("cherry")
print(fruits)
```

Output:

```
['kiwi', 'blueberry', 'banana', 'orange', 'strawberry', 'apple', 'mango']
```

Using pop() Function

The `pop()` function removes and returns the element at the specified position (index). If no index is specified, it removes and returns the last element.

```
# Removing element using pop() function
fruits.pop(3)
print(fruits)
```

Output:

```
['kiwi', 'blueberry', 'banana', 'strawberry', 'apple', 'mango']
```

Operations on Tuples

Similar to list, we can also perform the same operations on the tuples.

Creating a Tuple

You can create a tuple by placing all the elements inside parentheses (), separated by commas.

```
city = ("Delhi", "Madrid", "Chennai", "Rome")
```

Indexing in Tuple

Tuples are indexed similarly to lists, with both positive and negative indexing.

```
# Accessing elements using positive indexing
print(city[0])  # Output: Delhi
print(city[1])  # Output: Madrid

# Accessing elements using negative indexing
print(city[-1])  # Output: Rome
print(city[-2])  # Output: Chennai
```

Immutability

Tuples cannot be modified after they are created, which means their elements remain unchanged. Nevertheless, you can combine two or more tuples to form a new tuple.

The following line would generate an error

```
city[1] = "New York"
```

```
# Concatenating tuples
city2 = city + ("New York",)    #comma indicates that it is a tuple
print(city2)  # Output: city = ("Delhi", "Madrid", "Chennai", "Rome", "New York")
```

Dictionaries

A dictionary is a collection of key-value pairs. Dictionaries are defined by enclosing elements in curly braces { }.

Creating a Dictionary

A dictionary can be created by enclosing key-value pairs in curly braces {}, with each pair separated by commas. Within each pair, the key and value are divided by a colon ":"

```
# Creating a dictionary
car = {
    "make": "Toyota",
    "model": "Camry",
    "year": 2021,
    "colour": "Blue"
}
```

You can access the values in a dictionary by using the keys as shown:

```
print(car["make"])  # Output: Toyota
print(car["year"])   # Output: 2021
```

Mutability

In dictionaries, keys are immutable whereas values are mutable. You can change the values using the respective keys:

```
car["colour"]="Grey"
print(car)   #Ouput: {'make': 'Toyota', 'model': 'Camry', 'year': 2021, 'colour': 'Grey'}
You can also add key-value pairs in an existing dictionary:

car["mileage"] = 15000
print (car)  #Ouput: {'make': 'Toyota', 'model': 'Camry', 'year': 2021, 'colour':
'Grey', 'mileage': 15000}
```

Control Statements

Control statements in programming languages are essential tools that allow developers to dictate the flow of execution within their code. Depending on how the statements in a program will be executed, the flow of control in a program can be broadly classified into three types:

- Sequence Statements
- Conditional Statements
- Iterative Statements

Sequence Statement

In Python, a sequence statement refers to a set of instructions that are executed in the order they appear. This is the most straightforward form of control flow.

Program 12: To demonstrate the concept of Sequence statement

```
# Perform an Addition
num1 = 10
num2 = 20
result = num1 + num2
# Display the result
print("The result of", num1, "+", num2, "is:", result)
```

Output:

```
The result of 10 + 20 is: 30
```

Conditional Statements

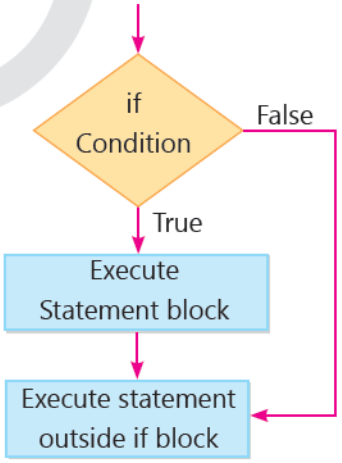
Conditional statement, also referred to as a selection statement, decision-making statement or branching statement, are used to control the flow of execution based on certain conditions. Conditional statements in Python allow you to execute certain blocks of code based on specific conditions.

In Python, there are various types of conditional statements, which are as follows:

- if statement
- if...else statement
- if...elif...else statement
- Nested if statement

The if Statement

The if statement is the simplest conditional statement. Here, a statement or a collection of statements within the if block are executed only if a certain condition or expression evaluates to True. If the condition evaluates to False, then the control of execution is passed to the next statement after the if block.



The syntax of the if statement is as follows:

```
if (Test Expression):
    Statement block
```

Program 13: To demonstrate the use of the if statement

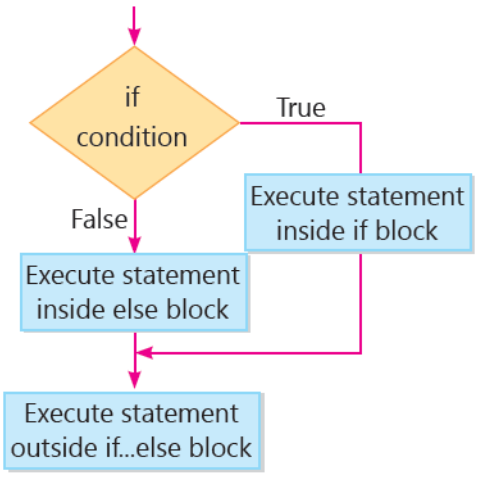
```
x = 10
if x > 5:
    print("x is greater than 5")
```

Output:

```
x is greater than 5
```

The if-else Statement

The if-else statement executes different blocks of code based on whether a condition is true or false. If the condition evaluates to True, the code block associated with the if statement is executed. If the condition evaluates to False, the code block associated with the else statement is executed.



The syntax of the if-else statement is as follows:

```
if (Test Expression):
    Statement block
else:
    Statement block
```

Program 14: To demonstrate the use of the if-else statement

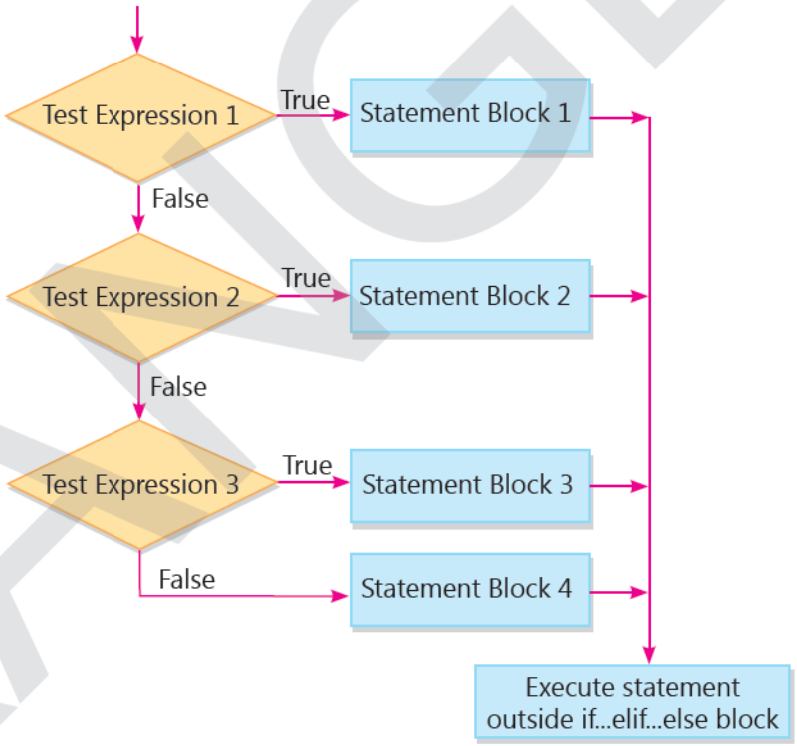
```
# Asking the user for input
temperature = float(input("Please enter the temperature in Celsius: "))
# Checking if the temperature is above or below freezing point
if temperature <= 0:
    print("It's freezing!")
else:
    print("It's not freezing.")
```

Output:

```
Please enter the temperature in Celsius: 35
It's not freezing.
```

The if-elif-else Statement

The if...elif...else statement helps us to test multiple conditions and follows a top-down approach. The if statement is used to check a single condition, while the elif (short for "else if") statement allows you to check additional conditions if the previous conditions are not met. If none of the conditions evaluates to True, then the final else statement gets executed. The if...elif...else statement provides a way to handle multiple decision branches in your code.



The syntax of the if...elif...else statement is as follows:

```
if (Test Expressions_1):
    Statement block 1
elif (Test Expression_2):
    Statement block 2
elif (Test Expression_3):
    Statement block 3
else:
    Statement block 4
```

Program 15: To demonstrate the use of the if-elif-else statement

```
# Asking the user for input
score = int( input("Please enter your exam score: "))
# Grading the score
if score >= 90:
    print("Your grade is A.")
elif score >= 80:
    print("Your grade is B.")
elif score >= 70:
    print("Your grade is C.")
```

```
elif score >= 60:
    print("Your grade is D.")
else:
    print("Your grade is F. You need to retake the exam.")
```

Output:

```
Please enter your exam score: 85
Your grade is B.
```

Nested If Statement

A nested if statement is a construct in programming where an if statement is placed within another if statement's block. This allows for the evaluation of multiple conditions in a hierarchical manner. In Python, indentation is crucial for defining the scope of each if statement.

The syntax of nested if statement is shown below:

```
if (Test Expression1):
    if (Test Expression2):
        Indented block 1
    else:
        Indented block 2
else:
    Indented block 3
```

Program 16: To demonstrate the use of the nested if statement

```
# Input the total purchase amount
total_purchase_amount = float(input("Enter the total purchase amount: Rs. "))

# Initialize the discount rate and discounted amount
discount_rate = 0
discounted_amount = 0

# Calculate the discount rate and discounted amount based on the total purchase
amount
if total_purchase_amount >= 10000:
    discount_rate = 0.10  # 10% discount
    print(f"Congratulations! You get a 10% discount.")
else:
    if total_purchase_amount >= 5000:
        discount_rate = 0.05  # 5% discount
        print(f"Congratulations! You get a 5% discount.")
    else:
      print("You are not eligible for getting discount!!")
# Calculate the discounted amount
discounted_amount = total_purchase_amount - (total_purchase_amount * discount_rate)
```

```
# Output the final amount after applying the discount
print(f"Total Amount: Rs. {total_purchase_amount:.2f}")
print(f"Discount Rate: {discount_rate * 100}%")
print(f"Discounted Amount: Rs.{discounted_amount:.2f}")
```

Output:

```
Enter the total purchase amount: Rs. 5500
Congratulations! You get a 5% discount.
Total Amount: Rs. 5500.00
Discount Rate: 5.0%
Discounted Amount: Rs. 5225.00
```

Ai Task

Write Python scripts for the following:

1. Accept 2 numbers from the user and display the larger number

2. Accept a number from the user. Check if the digit in the unit's place is odd or even. Example: if the user enters 452, 2 is the digit in the unit's place and is an even number.

Iterative Statements

Iterative statements are also known as looping statements. A loop is used to execute instructions or a block of code multiple times, without writing it repeatedly. A sequence of instructions when repeated for a number of times or until a condition is true is called a loop.

In Python, there are two types of conditional statements, which are as follows:

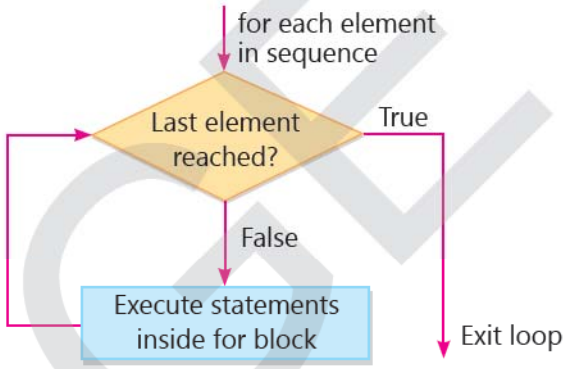
- The for loop
- The while loop

The for Loop

A for loop is used when you know in advance how many times you want to execute a block of code. It iterates over a sequence (such as a list, tuple, string, or range) and executes the block of code for each element in the sequence.

The for statement executes a simple or compound statement for a fixed number of times.

The syntax of while loop is as follows:




```
for element in sequence:
    statements
```

Program 17: To demonstrate the use of the for loop

```
numbers = [1, 2, 3, 4, 5]

for num in numbers:
    multiplied = num * 2
    print(multiplied)
```

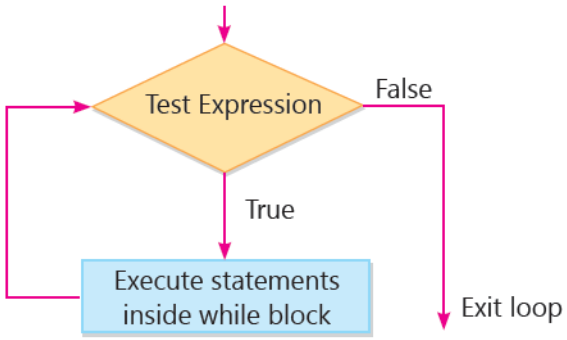
Output:

```
2
4
6
8
10
```

The while Loop

The while statement executes a set of statements repeatedly, until the logical expression evaluates to True. When the condition becomes False, the control comes out of the loop.

The syntax of while loop is as follows:




```
while (test expression):
    statements
    increment/decrement
```

Program 18: To demonstrate the use of the while loop

```
number = int(input("Enter a number to print its multiplication table: "))
counter = 1
print("Multiplication table of", number, ":")

while counter <= 10:
    print(number, "x", counter, "=", number * counter)
    counter += 1
```

Output:

```
Enter a number to print its multiplication table: 12
Multiplication table of 12:
12 x 1 = 12
12 x 2 = 24
12 x 3 = 36
12 x 4 = 48
12 x 5 = 60
12 x 6 = 72
12 x 7 = 84
12 x 8 = 96
12 x 9 = 108
12 x 10 = 120
```

Write Python scripts for the following:

1. Let us create a small guessing game using Python!

```
import random
print("Welcome to the Guessing Game!")
print("Try to guess the number I am thinking of between 1 and 20.")
num = random.randint(1, 20)  # generate random integer
my_guess = -1  # initialize guess to an invalid value
while my_guess != num:
    my_guess = int(input("Enter your guess: "))
    if my_guess < num:
        print("Too low, try again.")
```

2. Predict the output of the following code:

```
x = "I am eating mangos"

print("am" in x)
print("apple" in x)
```

3. Accept the 3 sides of a triangle. Check and display whether the triangle is Equilateral, Isosceles or Scalene [CBSE HANDBOOK]

The range() Function

The `range()` function in Python generates a sequence of numbers. It is commonly used with for loops to iterate over a specific range of numbers. The syntax of range function is as follows:

```
range(start, stop, step)
```

where,

- **start:** Starting number of the sequence.
- **stop:** Generate numbers up to, but not including last number.
- **step:** Difference between each number in the sequence.

Python use `range( )` function in three ways:

1. **range(stop):** This generates a sequence of numbers from 0 up to the specified stop value but not including it. It implicitly starts from 0 and increments by 1.

Program 19: To demonstrate the use of a range () with only stop value

```
for i in range(5):
    print(i)
```

Output:

```
0
1
2
3
4
```

2. **range(start, stop):** This generates a sequence of number beginning with start and going up to stop, but not including the specified stop value. It increments by 1 step.

Program 20: To demonstrate the use of a range () with start and stop values

```
for i in range(20, 25):
    print(i)
```

Output:

```
20
21
22
23
24
```

3. **range(start, stop, step):** This generates a sequence of number beginning with start and going up to stop, but not including the specified stop value. It allows you to specify the increment or decrement value by adding or subtracting the value.

Program 21: To demonstrate the use of a range () with start, stop, and step values

```
for i in range(25, 5, -5):
    print(i)
```

Output:

```
25
20
15
10
```

Ai Reboot

1. Fill in the blanks:
   a. The ..... statement is used to execute a block of code when the condition in if statement is False.
   b. The ..... loop executes continuously until the condition is False.
   c. The ..... function generates a sequence of numbers.
2. Give the output of the following code:

```
x = 5
while x > 0:
    print(x, end=" ")
    x -= 2
```

Jump Statements

Sometimes, there is a situation when the control of the program needs to be transferred out of the loop body, even if all the values of the iterations of the loop have not been completed. For this purpose, jumping statements are used in Python. Python offers different jumping statements—**break** and **continue**, which are used within the loop.

The break Statement

The **break** is a keyword in Python which is used for bringing the program control out of the loop. The **break** statement halts the execution of a loop and program flow switches to the statement after the loop. A single **break** statement will break out of only one loop.

The syntax of the break statement is as follows:

```
#loop statement
    break
```

Program 22: To demonstrate the use of the break statement using for loop

```
for number in range(0, 10):
    if(number == 5):
        break
    print('Number is:', number)
print('The loop exits!!!')
```

Output:

```
Number is: 0
Number is: 1
Number is: 2
Number is: 3
Number is: 4
The loop exits!!!
```

The continue Statement

The continue statement is used inside loops. When a continue statement is encountered inside a loop, then the control of the program jumps to the beginning of the loop for next iteration, skipping the execution of rest of the statements of the loop for the current iteration.

The syntax of the continue statement is as follows:

```
#loop statements
    continue
#the code to be skipped
```

Program 23: To demonstrate the use of the continue statement

```
for i in range(1, 6):
    if i == 3:
        continue  # Skip the rest of the loop body for i == 3
    print("Iteration:", i)
```

Output:

```
Iteration: 1
Iteration: 2
Iteration: 4
Iteration: 5
```

Digital Literacy

Video Session

Watch the video Python Lists, Tuples And Dictionaries - 10 | Python For Beginners | Python Tutorial | Simplilearn - https://www.youtube.com/watch?v=fXRxHrDhQuI

Ai Task

Write Python programs for the following:

1. Write a Python program to accept a string. Display the length of each word in the string.

2. Display all even numbers between 20 and 80.

3. Assume a list of 10 numbers. Display:
   a. Count of even numbers
   b. Count of odd numbers
   c. Sum of all even numbers
   d. Sum of all odd numbers
   e. Count of all numbers divisible by 10

4. Create a dictionary named student with keys "name", "age", and "grade", and assign appropriate values to them. Now write Python code to do the following:
   a. Change the value of grade from "A" to "A+"
   b. Display the value of age
   c. Add another key-value pair to the dictionary – "subject" : "Computer Science"
   d. Display the dictionary

CSV Files

A CSV (Comma-Separated Values) file is a popular and widely used format for storing and exchanging tabular data. It serves as a lightweight and versatile means of organising data into rows and columns, akin to a spreadsheet or a database table. In its simplest form, each line of a CSV file represents a single record, with individual data fields separated by commas. However, CSV files offer flexibility in delimiters, allowing for the use of alternative separators like semicolons or tabs. Typically, the first row of a CSV file contains headers, defining the names of each column, while subsequent rows hold the corresponding data entries. Due to its text-based nature, CSV files are platform-independent and easily readable by both humans and machines. This simplicity, combined with its compatibility with various programming languages and tools, makes CSV a go-to choice for data interchange, storage, and analysis across diverse domains such as finance, research, and software development. There are various ways by which you can create a CSV file, such as Excel (by saving a spreadsheet with .csv extension).

For example, the CSV file contains the following data:

```
Customer_ID, First_Name, Last_Name, City
1, Akash, Patel, Mumbai
2, Priya, Sharma, Delhi
3, Aarav, Singh, Jaipur
4, Neha, Trivedi, Ahmedabad
5, Rahul, Mehta, Bangalore
```

Some basic operations of CSV files are as follows:

- **Importing the CSV library:** In Python, the csv module provides functionality to work with CSV files. It includes classes to read and write tabular data in CSV format.

```
import csv
```

- **Opening a CSV file in reading mode:** When opening a file in reading mode ('r'), you are telling Python that you only intend to read from the file, not modify it. The `csv.reader()` function then reads the contents of the file line by line. Note that, you have already created CSV file that you want to open. For example, Customer.csv.

Program 24: To opening a CSV file in reading mode

```
import csv
with open('Customer.csv', 'r') as file:
    reader = csv.reader(file)
    for row in reader:
        # display data
```

- **Opening a CSV File in writing mode:** When opening a file in writing mode ('w'), you are telling Python that you intend to write to the file. If the file already exists, it will be truncated (emptied) first. If it doesn't exist, a new file will be created.

Program 25: To opening a CSV file in writing mode

```
import csv
with open('Customer.csv', 'w', newline='') as file:
    writer = csv.writer(file)
    # Write rows here
```

- **Closing a CSV file:** In Python, it is important to close files after you have finished working with them. However, using the with statement automatically closes the file when the block is exited, so you don't have to worry about explicitly closing it.

- **Writing rows to a CSV file:** You can use the `writerows()` method to write multiple rows to a CSV file at once. Each row should be a list of values. This function will replace all existing data in the CSV file.

Program 26: To write data to a CSV file row by row

```
import csv
# Open the CSV file in write mode
with open("books.csv", mode='w', newline='') as file:
    writer = csv.writer(file)
    # Write the header row
    writer.writerow(["Title", "Author", "Year", "Price"])
    while True:
        # Accept book details from the user
        title = input("Enter the book title (or type 'exit' to finish): ")
        if title.lower() == 'exit':
            break
        author = input("Enter the author's name: ")
        year = input("Enter the publication year: ")
        price = input("Enter the price: ")
        # Write the book details to the CSV file
        writer.writerow([title, author, year, price])
print("Details of books have been written to the csv file")
```

Output:

```
Enter the book title (or type 'exit' to finish): Harry Potter
Enter the author's name: J K Rowling
Enter the publication year: 2008
Enter the price: 345
Book details have been written to the CSV file.
Enter the book title (or type 'exit' to finish): Twilight
Enter the author's name: S Mayer
Enter the publication year: 2010
Enter the price: 458
Book details have been written to the CSV file.
Enter the book title (or type 'exit' to finish): exit
Details of books have been written to the csv file
```

- **Append rows to a CSV file:** Appending a row to an existing CSV file involves opening the file in append mode ('a'), and writing a single row of data to it. You can achieve this using the csv.writer object's `writerow()` method.

Program 27: To write multiple rows in one go to a CSV file.

```
import csv
data =[['4', 'Neha', 'Trivedi', 'Ahmedabad'],
['5', 'Rahul', 'Mehta', 'Bangalore']]

with open('Customer.csv', 'a', newline='') as file:
    writer = csv.writer(file)
```

```
    writer.writerows(data)

with open('Customer.csv', 'r') as file:
    reader = csv.reader(file)
    for row in reader:
        print(row)
```

Output:

```
['Customer_ID', 'First_Name', 'Last_Name', 'City']
['1', 'Aishwarya', 'Iyer', 'Chennai']
['2', 'Rajesh', 'Reddy', 'Hyderabad']
['3', 'Sneha', 'Gupta', 'Kolkata']
['4', 'Neha', 'Trivedi', 'Ahmedabad']
['5', 'Rahul', 'Mehta', 'Bangalore']
```

Brainy Fact

In artificial intelligence and machine learning, CSV files are frequently used to store datasets for training models. The data is typically arranged in rows and columns, with each column representing a variable or feature and each row representing an instance or example.

Understanding Libraries

In Python, libraries refer to collections of pre-written code modules or functions that you can use to perform specific tasks without having to write the code from scratch. These libraries are designed to be reused in different projects that makes the development easier by providing various ready-to-use solutions for particular tasks.

In Python, functions are organised within libraries similar to how library stores books on different subjects. Python libraries organise functions into different categories such as, math, web development, data analysis, machine learning, etc. These libraries can be imported into Python programs to access their functionalities and tools.

For example, if you want to use a specific function like `sqrt()`, `pow()`, `abs()`, or `sin()` in Python, you have to tell Python that you want to use the math library where this function is stored. You do this by adding the line 'import math' at the beginning of your program. This lets you access and use all the helpful tools that the math library offers, including the '`sqrt()`' function.

Python offers a wide range of libraries such as, NumPy, Pandas, Matpoltlib, Scikit-Learn, etc. for various purposes that makes Python a versatile language. With these libraries, you can do the various tasks, such as, web development, data analysis, machine learning, scientific computing, etc.

Introduction to NumPy

NumPy is the short form of Numerical Python. It is a fundamental library in Python that is used for performing numerical computation. It provides support for arrays, matrices, and a variety of mathematical functions to operate on these data structures efficiently. Its array-based data structures and operations execution makes it very useful for various applications, such as data analysis, machine learning, scientific computing, etc.

In NumPy, there are several types of arrays, which are as follows:

- **One-dimensional Arrays (1D Arrays):** These arrays contain elements arranged in a single row or column. One-dimensional arrays are created using the `numpy.array()` function with a Python list or tuple as input.

Program 28: To demonstrate the use of 1D array

```
import numpy as np
arr_1d = np.array([1, 2, 3, 4, 5])
print(arr_1d)
```

Output:

```
[1 2 3 4 5]
```

- **Two-dimensional Arrays (2D Arrays):** Two-dimensional arrays are arranged in rows and columns, forming a grid-like structure. Two-dimensional arrays are created using nested lists or by reshaping a one-dimensional array.

Program 29: To demonstrate the use of 2D array

```
import numpy as np
arr_2d = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])
print(arr_2d)
```

Output:

```
[[1 2 3]
 [4 5 6]
 [7 8 9]]
```

- **Multi-dimensional Arrays (nD Arrays):** These arrays have more than two dimensions, which helps in complex data representations. Multi-dimensional arrays can be created using nested lists or by reshaping existing arrays.

Program 30: To demonstrate the use of nD array

```
import numpy as np
# Creating a 3x3x3 ndarray
arr_nd = np.array([[[1, 2, 3], [4, 5, 6], [7, 8, 9]],
                   [[10, 11, 12], [13, 14, 15], [16, 17, 18]],
                   [[19, 20, 21], [22, 23, 24], [25, 26, 27]]])
print(arr_nd)
```

Output:

```
[[[ 1  2  3]
  [ 4  5  6]
  [ 7  8  9]]

 [[10 11 12]
  [13 14 15]
  [16 17 18]]

 [[19 20 21]
  [22 23 24]
  [25 26 27]]]
```

In NumPy, arrays are homogeneous, it means all elements in an array must be of the same data type, for example, integers, floats, etc.

You can install NumPy using pip. For installing NumPy, you need to open your terminal or command prompt and run the following command:

```
pip install numpy
```

NumPy Library in Artificial Intelligence

Let us understand why and where we can use the NumPy library in Artificial Intelligence with the help of an example.

Suppose, you have a dataset containing daily temperature readings from weather stations across different cities. You can utilise NumPy arrays to efficiently manage and analyse this data.

With NumPy's array operations, you can easily perform the following tasks:

- Calculating the average temperature for each city over the recorded days.
- Finding the total temperature recorded for each day across all cities.
- Determining the overall average temperature across all cities and days.
- Identifying the highest and lowest temperatures recorded.

NumPy's array operations streamline these computations that enables you to handle large datasets with ease. This makes NumPy an indispensable tool for processing and analysing data in different fields.

Creating a NumPy Array

NumPy array are created in several ways:

1. Using `np.array()` to create arrays from lists or tuples

Program 31: To demonstrate the use of np.array() to create arrays from lists or tuples

```
import numpy as np
# From a list
arr1 = np.array([1, 2, 3, 4, 5])
print(arr1)
# From a tuple
arr2 = np.array((6, 7, 8, 9, 10))
print(arr2)
```


Output:

```
[1 2 3 4 5]
[ 6  7  8  9 10]
```

2. Using `np.zeros()` to create an array of zeros

The `zeros()` creates arrays filled with zeros.

Program 32: To demonstrate the use of np.zeros() to create an array of zeros

```
import numpy as np
# Array of zeros
arr_zeros = np.zeros(5)
print(arr_zeros)
```

Output:

```
[0. 0. 0. 0. 0.]
```

3. Using `np.ones()` to create an array of ones

The `ones()` function creates arrays filled with ones.

Program 33: To demonstrate the use of np.ones() to create an array of ones

```
import numpy as np
# Array of ones
arr_ones = np.ones(5)
print(arr_ones)
```

Output:

```
[1. 1. 1. 1. 1.]
```

4. Using `arange()` to create an array

The `arange()` function in NumPy is used to create arrays with regularly incrementing values. The syntax of `arange()` function is as follows:

```
numpy.arange(start, stop, step, dtype=None)
```

where,

- **start:** The start of the interval (inclusive). If not specified, defaults to 0.
- **stop:** The end of the interval (exclusive).
- **step:** The spacing between values. If not specified, defaults to 1.
- **dtype:** The data type of the array. If not specified, NumPy will infer the data type.

Program 34: To demonstrate the use of arange() to create an array

```
import numpy as np
# Array with a range of values
arr_range = np.arange(0, 10, 2)
print(arr_range)
```

Output:

```
[0 2 4 6 8]
```

5. Create an array in NumPy using `np.empty()`

The `np.empty()` function in NumPy is used to create a new array of specified shape and data type, without initialising the elements. This means that the values in the array are not set and can be any random values that were already in the memory location used for the array.

Program 35: To demonstrate the use of np.empty() to create an array

```
import numpy as np
# Create an empty array
arr_empty = np.empty((3, 3))  # Creating a 3x3 empty array
print(arr_empty)
```

Output:

```
[[5.4e-323 0.0e+000 0.0e+000]
[0.0e+000 6.4e-323 0.0e+000]
[0.0e+000 0.0e+000 3.0e-323]]
```

Program 36: To create a NumPy array by taking values from the user through np.empty()

```
import numpy as np

# Get the dimensions of the array from the user
rows = int(input("Enter the number of rows: "))
cols = int(input("Enter the number of columns: "))

# Create an empty array
array = np.empty((rows, cols))

# Get the values from the user and fill the array
for i in range(rows):
    for j in range(cols):
        value = float(input(f"Enter the value for element ({i},{j}): "))
        array[i, j] = value

# Print the array
print("Array:")
print(array)
```

Output:

```
Enter the number of rows: 2
Enter the number of columns: 3
Enter the value for element (0,0): 25
Enter the value for element (0,1): 36
Enter the value for element (0,2): 25
Enter the value for element (1,0): 75
Enter the value for element (1,1): 86
Enter the value for element (1,2): 35
```

```
Array:
[[25. 36. 25.]
[75. 86. 35.]]
```

Program 37: To create a numpy array by taking values from the user through np.empty()

```
import numpy as np

# Get the dimensions of the array from the user
rows = int(input("Enter the number of rows: "))
cols = int(input("Enter the number of columns: "))

# Create an empty array
array = np.empty((rows, cols))

# Get the values from the user and fill the array
for i in range(rows):
    for j in range(cols):
        value = float(input(f"Enter the value for element ({i},{j}): "))
        array[i, j] = value

# Print the array
print("Array:")
print(array)
```

Output:

```
Enter the number of rows: 2
Enter the number of columns: 3
Enter the value for element (0,0): 25
Enter the value for element (0,1): 36
Enter the value for element (0,2): 25
Enter the value for element (1,0): 75
Enter the value for element (1,1): 86
Enter the value for element (1,2): 35
Array:
[[25. 36. 25.]
 [75. 86. 35.]]
```

Introduction to Pandas

Pandas is a popular Python library widely used for data manipulation and analysis. The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis". It provides data structures and functions that make it easy to work with structured data, such as tabular data (for example Excel spreadsheets or SQL tables). Pandas is built on top of NumPy, another Python library for numerical computing, and it extends its functionality by providing high-level data structures and powerful tools for data manipulation, cleaning, filtering, grouping, merging, etc. Its adaptability and user-friendly interface make this an essential tool for data analysts, scientists, and engineers who are working on the structured data.

You can install Pandas using pip. For installing Pandas, you need to open your terminal or command prompt and run the following command:

```
pip install pandas
```

Pandas Library in Artificial Intelligence

The Pandas library plays a crucial role in Artificial Intelligence and data science workflows. It provides powerful and flexible tools for data manipulation and analysis, which are essential for preparing and exploring datasets used in AI models.

Let us understand why and where we can use the Pandas library in Artificial Intelligence with the help of an example.

Consider a dataset containing information about student performance, including grades, attendance, extracurricular activities, and demographic details. Pandas can be leveraged to import the dataset, compute statistical summaries, and conduct in-depth analyses to understand the factors influencing academic success.

Pandas' powerful data manipulation capabilities enable educators and administrators to identify trends, correlations, and patterns within the student body. By examining variables such as attendance rates, participation in extracurricular activities, and socio-economic backgrounds, schools can gain insights into factors affecting student achievement and tailor interventions to support student success.

Pandas provides powerful capabilities for data manipulation and aggregation by simplifying the execution of complex analyses. These capabilities play a vital role in AI and data-driven decision-making, allowing businesses to derive actionable insights from their data with ease.

Data Structure of Pandas

Pandas provides two primary data structures: Series and DataFrame, which are used to store and manipulate data efficiently.

Series

Series represents a one-dimensional labeled array capable of holding various datatypes in a series (integer, float, string, Python objects, etc.). Each element in a series has an index label, which can be used to access the elements. By default, Series have numeric data labels starting from zero, with each value associated with an index label.

Index	Data
0	Class 12
1	Class 11
2	Class 10
3	Class 9
4	Class 8

Brainy Fact

Pandas performs better with datasets containing 500,000 rows or more, while NumPy excels with datasets of 50,000 rows or fewer. When it comes to indexing, Pandas series are significantly slower than NumPy arrays, which have very fast indexing capabilities.

Program 38: To create a series from a list

```
import pandas as pd
# Creating a series from a list
data = [1, 2, 3, 4, 5]
series = pd.Series(data)
print(series)
```

Output:

```
0    1
1    2
2    3
3    4
4    5
dtype: int64
```

DataFrames

In pandas, a DataFrame is a two-dimensional labeled data structure, similar to a table in a spreadsheet or a database. It consists of rows and columns, where each column can have a different data type (like integers, floats, strings, etc.). Some example of a DataFrame are: class's result, menu items in a restaurant, or a train's reservation chart, etc.

Index	Name	Test 1	Test 2	Test 3
0	Amit	92	75	79
1	Yash	81	85	84
2	Rohan	76	72	85
3	Devesh	78	83	90
4	Nihar	92	87	87

Creating a DataFrame

There are several ways depending on the data source and structure by which you can create a dataframe in Pandas. Some most common method for creating a DataFrame are:

- **From Lists or Arrays:** You can create a DataFrame by passing a list or a NumPy array to the `pd.DataFrame()` constructor. Each element of the list or array will be treated as a row.

Program 39: To create a DataFrame using lists or arrays

```
import pandas as pd
data = [[1, 'Mayur', 25],
```

```
        [2, 'Esha', 30],
        [3, 'Vedant', 28]]
df = pd.DataFrame(data, columns=['Name', 'Age', 'City'])
print(df)
```

Output:

```
     Name       Age    City
0       1     Mayur      25
1       2      Esha      30
2       3    Vedant      28
```

- **From a Dictionary of Lists or Arrays:** You can create a DataFrame from a dictionary where keys are column names and values are lists or arrays containing the data for each column.

Program 40: To create a DataFrame using dictionary of lists or arrays

```
import pandas as pd
data = {'Name': ['Yash', 'Rehan', 'Neha'],
        'Age': [25, 30, 28],
        'City': ['Delhi', 'Mumbai', 'Bengaluru']}
df = pd.DataFrame(data)
print(df)
```

Output:

```
    Name  Age       City
0   Yash   25      Delhi
1  Rehan   30     Mumbai
2   Neha   28  Bengaluru
```

- **Using NumPy ndarrays:** Creating DataFrames from NumPy ndarrays allows you to leverage the power of both libraries for efficient data processing and analysis.

Program 41: To create a DataFrame using ndarray

```
import pandas as pd
import numpy as np
# Using NumPy ndarray
data = np.array([[1, 'Mayur', 25],
                 [2, 'Esha', 30],
                 [3, 'Vedant', 28]])
df = pd.DataFrame(data, columns=['ID', 'Name', 'Age'])
print(df)
```

Output:

```
    ID       Name     Age
0    1      Mayur      25
1    2       Esha      30
2    3     Vedant      28
```

- **Using a List of Dictionaries:** You can create a DataFrame from a list of dictionaries where each dictionary represents a row and keys represent column names.

Program 42: To create a DataFrame using ndarray

```
import pandas as pd
# Using a List of Dictionaries
data = [{'ID': 1, 'Name': 'Arti', 'Age': 25},
        {'ID': 2, 'Name': 'Trinabh', 'Age': 30},
        {'ID': 3, 'Name': 'Surbhi', 'Age': 28}]
df = pd.DataFrame(data)
print(df)
```

Output:

```
    ID      Name     Age
0    1      Arti      25
1    2   Trinabh      30
2    3    Surbhi      28
```

Rows and Columns in DataFrame

A DataFrame is a core data structure in Pandas, providing a tabular representation of data with rows and columns. This two-dimensional format mirrors the structure of a typical spreadsheet or a database table. Each column in a DataFrame represents a different variable or feature, while each row corresponds to a single observation or data point or record.

You can perform some fundamental operations on rows and columns, which are as follows:

- **Selection:** DataFrames allow us to select specific rows or columns based on various criteria. This selection can be done using labels (column names) or indices (row numbers) or through boolean indexing based on conditions.
- **Addition:** New columns or rows can be added to a DataFrame to incorporate additional information or derived features. These additions can be based on existing data within the DataFrame or computed using external sources.
- **Deletion:** DataFrames enable the removal of rows or columns that are not needed for analysis. This process helps in cleaning and preparing the data for further processing.

To perform these operation on dataframe, we need to first create a dataframe.

Program 43: To create a sample DataFrame

```
# Import pandas library
import pandas as pd
# Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}
# Convert the dictionary into DataFrame
df = pd.DataFrame(data)
# Displaying DataFrame
print(df)
```

Output:

```
      Name  Age  Address Qualification
0     Adit   27    Delhi           Msc
1     Ekam   24   Kanpur            MA
2   Sakshi   25   Meerut           MCA
3      Anu   30   Indore           PhD
```

Selecting Column from the DataFrame

You can select specific columns from a DataFrame using their labels (column names). This allows you to focus on relevant data for analysis.

Program 44: To select a column from the DataFrame

```
# Import pandas library
import pandas as pd
# Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}
# Convert the dictionary into DataFrame
df = pd.DataFrame(data)
# Displaying DataFrame with Selected Column
print(df[['Name', 'Age', 'Qualification']])
```

Output:

```
      Name     Age   Qualification
0     Adit      27             Msc
1     Ekam      24              MA
2   Sakshi      25             MCA
3      Anu      30             PhD
```

Selecting Row from the DataFrame

You can select specific rows based on their indices or using boolean indexing based on conditions. This enables you to filter the data based on certain criteria, such as values meeting a particular condition. To select a particular row from the DataFrame, you can use the .iloc[] method. This method is used for integer-location based indexing, which means you select rows and columns by their position (index).

Program 45: To select row using .iloc[] method

```
# Import pandas library
import pandas as pd

# Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
```


```
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}

# Convert the dictionary into DataFrame
df = pd.DataFrame(data)

# Selecting the second row (index 1)
selected_row = df.iloc[2]
print(selected_row)
```

Output:

```
Name              Sakshi
Age                   25
Address           Meerut
Qualification        MCA
Name: 2, dtype: object
```

To print the output in the same format as the original DataFrame (df), you can use the to_dict() method along with orient='records' to convert the selected row into a dictionary and then create a new DataFrame from that dictionary.

Program 46: To select a row using .iloc[] method and print the output in the same format as the original DataFrame

```
# Import pandas library
import pandas as pd

# Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}

# Convert the dictionary into DataFrame
df = pd.DataFrame(data)

# Selecting the second row (index 1)
selected_row = df.iloc[2]

# Convert the selected row to a dictionary
selected_row_dict = selected_row.to_dict()

# Create a new DataFrame from the selected row dictionary
selected_row_df = pd.DataFrame([selected_row_dict])

# Display the new DataFrame
print(selected_row_df)
```

Output:

```
      Name    Age  Address   Qualification
0   Sakshi     25   Meerut             MCA
```

Selecting the Specific Element

To access a specific element in a DataFrame, you can use the iloc[] method to access the element by its row and column index.

Program 47: To select a specific element from the DataFrame

```
# Import pandas library
import pandas as pd

# Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}

# Convert the dictionary into DataFrame
df = pd.DataFrame(data)

# Accessing a specific element
element = df.iloc[1, 2]  # Row index 1, Column index 2
print("Element at (1, 2):", element)
```

Output:

```
Element at (1, 2): Kanpur
```

Adding a New Column to the DataFrame

You can add a column in a DataFrame by directly assigning values to a new column using the DataFrame's bracket notation. This will add the columns at the end of the DataFrame.

Program 48: To add a new column to the end of the DataFrame

```
# Import pandas library
import pandas as pd

# Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}

# Convert the dictionary into DataFrame
df = pd.DataFrame(data)
```


```
# Add a new column 'Salary' to the DataFrame
df['Salary'] = [50000, 45000, 55000, 60000]

# Displaying DataFrame
    print(df[['Name', 'Age', 'Address', 'Qualification', 'Salary']])
```

Output:

```
      Name   Age  Address  Qualification    Salary
0     Adit    27    Delhi            Msc     50000
1     Ekam    24   Kanpur             MA     45000
2   Sakshi    25   Meerut            MCA     55000
3      Anu    30   Indore            PhD     60000
```

If you want to add the column at specific position in the DataFrame, then you need to use the insert() function that allows you to specify the index value where you want to insert the column.

Program 49: To add the column at specific position in the DataFrame

```
# Import pandas library
import pandas as pd

 # Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}

# Convert the dictionary into DataFrame
df = pd.DataFrame(data)

# Specifying values to new column
Salary = [50000, 45000, 55000, 60000]

# Index position where you want to insert the new column (zero-based)
index_position = 2  # Inserting the new column after the 'Age' column

# Insert new column at specified index position
df.insert(index_position, 'Salary', Salary)

# Displaying DataFrame
print(df)
```

Output:

```
      Name   Age   Salary  Address   Qualification
0     Adit    27    50000    Delhi             Msc
1     Ekam    24    45000   Kanpur              MA
2   Sakshi    25    55000   Meerut             MCA
3      Anu    30    60000   Indore             PhD
```

Adding a New Row to the DataFrame

A new row can be added to a DataFrame by utilising the DataFrame.loc[] method. This method allows you to add rows by specifying the index and the values.

Program 50: To add the row at the end in the DataFrame

```
# Import pandas library
import pandas as pd

# Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}
# Convert the dictionary into DataFrame
df = pd.DataFrame(data)

# Add a new column 'Salary' to the DataFrame
df['Salary'] = [50000, 45000, 55000, 60000]

# Add a new row to the DataFrame using DataFrame.loc[]
new_row = {'Name': 'Himanshi', 'Age': 29, 'Address': 'Mumbai', 'Qualification':
'B.Tech', 'Salary': 52000}
df.loc[len(df)] = new_row

# Display the DataFrame with the new row
print("DataFrame with the new row:")
print(df)
```

Output:

```
DataFrame with the new row:
       Name   Age  Address Qualification   Salary
0      Adit    27    Delhi           Msc    50000
1      Ekam    24   Kanpur            MA    45000
2    Sakshi    25   Meerut           MCA    55000
3       Anu    30   Indore           PhD    60000
4  Himanshi    29   Mumbai       B.Tech    52000
```

To add a new row in the middle of the DataFrame, you can use the DataFrame.iloc[] method to specify the index where you want to insert the row.

Program 51: To add the row at middle in the DataFrame

```
# Import pandas library
import pandas as pd
```

```
# Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}

# Convert the dictionary into DataFrame
df = pd.DataFrame(data)

# Add a new column 'Salary' to the DataFrame
df['Salary'] = [50000, 45000, 55000, 60000]

# Add a new row to the DataFrame using DataFrame.loc[]
new_row = {'Name': 'Himanshi', 'Age': 29, 'Address': 'Mumbai', 'Qualification':
'B.Tech', 'Salary': 52000}

# Index where you want to insert the row
index_to_insert = len(df)//2
# Shift existing rows down and insert new row at index_to_insert
df = pd.concat([df.iloc[:index_to_insert], pd.DataFrame([new_row]), df.iloc[index_
to_insert:]]).reset_index(drop=True)

# Display the DataFrame with the new row in the middle
print("DataFrame with the new row in the middle:")
print(df)
```

Output:

```
DataFrame with the new row in the middle:
       Name  Age Address Qualification  Salary
0      Adit   27   Delhi           Msc   50000
1      Ekam   24  Kanpur            MA   45000
2  Himanshi   29  Mumbai         B.Tech   52000
3    Sakshi   25  Meerut            MCA   55000
4       Anu   30  Indore            PhD   60000
```

Deleting Rows & Columns from a DataFrame

Deleting rows and columns from a DataFrame can be accomplished using the drop() method. This method allows you to specify which rows or columns to delete and whether to make the changes in place or return a new DataFrame.

To remove rows and columns from a DataFrame, we specify the labels' names and the axis (0 for rows, 1 for columns).

Program 52: To delete rows and columns from the DataFrame

```
# Import pandas library
import pandas as pd
```

```
# Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}

# Convert the dictionary into DataFrame
df = pd.DataFrame(data)

# Add a new column 'Salary' to the DataFrame
df['Salary'] = [50000, 45000, 55000, 60000]

# Deleting a Row
df.drop(index=1, inplace=True)  # Deleting the row with index 1 (Ekam)
print("\nDataFrame after deleting a row:")
print(df)

# Deleting a Column
df.drop(columns='Address', inplace=True)  # Deleting the 'Address' column

print("DataFrame after deleting a column:")
print(df)
```

Output:

```
DataFrame after deleting a row:
      Name   Age  Address Qualification   Salary
0     Adit    27    Delhi           Msc    50000
2   Sakshi    25   Meerut           MCA    55000
3      Anu    30   Indore           PhD    60000

DataFrame after deleting a column:
      Name   Age  Qualification  Salary
0     Adit    27            Msc   50000
2   Sakshi    25            MCA   55000
3      Anu    30            PhD   60000
```

To delete multiple columns with specific labels, you can modify the drop method to include a list of column labels. `DataFrame.drop()` method can also be used to remove the duplicate rows.

```
# Deleting Columns with specific labels
labels_to_delete = ['Age', 'Address'] #specify in a list
df.drop(columns=labels_to_delete, inplace=True)
```


AI Reboot

1. Fill in the blanks:
   a. The two data structures that are supported by Pandas are ..... and ..... .
   b. The ..... library in Python excels in creating N-dimension data objects.
   c. The statement to install NumPy is ..... .
   d. You can check the shape of an array by using the ..... method in NumPy.
2. Answer the following questions
   a. What is a DataFrame in Pandas?

      .....

   b. Give one advantage of using NumPy arrays over lists.

      .....

   c. We can create a data frame from a CSV file. (State True or False) .....

Understanding Missing Values

Understanding missing values in a DataFrame is crucial for data analysis, cleaning and preprocessing. Missing values, often denoted as NaN (Not a Number) or None, can occur due to various reasons such as data entry errors, incomplete data, or data transformation processes. For example, while reviews for a product online, some customers may not provide feedback on every aspect of the product if they did not use all of its features. Dealing with missing values is an essential step in data cleaning and preprocessing.

The `isnull()` function in Pandas is used to detect missing or NaN (Not a Number) values within a DataFrame. It returns a DataFrame of the same shape as the original DataFrame, where each element is a boolean value indicating whether it's missing (True) or not (False). The `isnull()` function returns True for missing values and False for non-missing values.

Employee Name	Salary
Rohit	50000
Pankaj	52000
Vivan	NaN
Chirag	53000
Sanjay	NaN

`isnull()` →

Employee Name	Salary
False	False
False	False
False	True
False	False
False	True

You can count the number of missing values in each column or row.

- **df.isnull().sum():** Returns the number of missing values in each column.
- **df.isnull().sum(axis=1):** Returns the number of missing values in each row.

There are several ways by which you can handle missing values in DataFrame:

- **Imputation:** Imputation involves replacing missing values with a specific value. Common strategies include replacing missing values with the mean, median, or mode of the column. This method helps in retaining the structure of the dataset and avoids losing valuable information.

  Pandas provides methods like `fillna()` to perform imputation. For example, you can fill missing values in a DataFrame `df` with the mean of each column using `df.fillna(df.mean())`.

- **Dropping:** Dropping involves removing rows or columns containing missing values. This method is useful when the missing values are sparse and dropping them doesn't significantly impact the analysis.

  Pandas provides the `dropna()` function to drop rows or columns with missing values. For example, you can drop rows with any missing values in a DataFrame `df` using `df.dropna()`.

- **Interpolation:** Interpolation involves estimating missing values based on existing data. Pandas provides interpolation methods such as `interpolate()` to estimate missing values. For example, you can perform linear interpolation on a DataFrame `df` using `df.interpolate()`. This method is particularly useful for time series or ordered data where missing values can be inferred from neighbouring values.

Program 53: To fill the missing value in the DataFrame

```
import pandas as pd
import numpy as np
# Define a dictionary containing data with some missing values
data = {
    'Name': ['Adit', 'Ekam', 'Sakshi', 'Anu'],
    'Age': [27, np.nan, 25, 30],
    'Address': ['Delhi', 'Kanpur', np.nan, 'Indore'],
    'Qualification': ['MSc', 'MA', 'MCA', 'PhD']
}
# Convert the dictionary into a DataFrame
df = pd.DataFrame(data)
# Display the original DataFrame
print("Original DataFrame:")
print(df)
# Finding any missing value in a column
print("\nMissing values in each column:")
print(df.isnull().sum())
# Finding the total number of NaN values
print("\nTotal number of NaN values:")
print(df.isnull().sum().sum())
# Deleting entire row with NaN values
df_dropped = df.dropna()
print("\nDataFrame after dropping rows with NaN values:")
print(df_dropped)
# Filling NaN values with mean in 'Age' column and 'Chennai' in 'Address' column
# The mean value is rounded to an integer
df_filled = df.fillna({'Age': round(df['Age'].mean()), 'Address': 'Chennai'})
print("\nDataFrame after filling NaN values:")
print(df_filled)
```


Output:

```
Original DataFrame:
     Name   Age Address Qualification
0    Adit  27.0   Delhi           MSc
1    Ekam   NaN  Kanpur            MA
2  Sakshi  25.0     NaN           MCA
3     Anu  30.0  Indore           PhD

Missing values in each column:
Name             0
Age              1
Address          1
Qualification    0
dtype: int64

Total number of NaN values:
2

DataFrame after dropping rows with NaN values:
   Name   Age Address Qualification
0  Adit  27.0   Delhi           MSc
3   Anu  30.0  Indore           PhD

DataFrame after filling NaN values:
     Name   Age  Address Qualification
0    Adit  27.0    Delhi           MSc
1    Ekam  27.0   Kanpur            MA
2  Sakshi  25.0  Chennai           MCA
3     Anu  30.0   Indore           PhD
```

Attributes of DataFrames

In Pandas, attributes are properties that describe various aspects of the DataFrame's structure and content. Attributes do not perform any computation or modification on the data; instead, they provide access to metadata or properties of the object. They are accessed using dot notation and do not require parentheses to invoke. The syntax for using an attribute is:

```
DataFrame_name.attribute
```

Program 54: To create sample DataFrame to illustrate its common attributes and methods

```
import pandas as pd
# Create a DataFrame with some product data
data = {
    'Product': ['Laptop', 'Tablet', 'Smartphone', 'Monitor'],
    'Price': [1000, 500, 800, 300],
    'Stock': [50, 150, 200, 100],
```

```
    'Rating': [4.5, 4.0, 4.7, 4.3]
}
df = pd.DataFrame(data)
# Display the DataFrame
print(df)
```

Output:

```
      Product  Price  Stock  Rating
0      Laptop   1000     50     4.5
1      Tablet    500    150     4.0
2  Smartphone    800    200     4.7
3     Monitor    300    100     4.3
```

Some common attributes of DataFrame are as follows:

- **columns:** This attribute returns an Index object containing the column labels of the DataFrame.

  ```
  print(df.columns)
  ```

  Output: `Index(['Product', 'Price', 'Stock', 'Rating'], dtype='object')`

- **index:** This attribute returns the row labels of the DataFrame.

  ```
  print(df.index)
  ```

  Output: `RangeIndex(start=0, stop=4, step=1)`

- **dtypes:** This attribute returns the data types of each column in the DataFrame.

  ```
  print(df.dtypes)
  ```

  Output:

  ```
  Product     object
  Price        int64
  Stock        int64
  Rating     float64
  dtype: object
  ```

- **shape:** This attribute returns a tuple representing the dimensions of the DataFrame, i.e., (number of rows, number of columns).

  ```
  print(df.shape)
  ```

  Output: `(4, 4)`

- **size:** This attribute returns the total number of elements in the DataFrame.

  ```
  print(df.size)
  ```

  Output: `16`

- **values:** This attribute returns a two-dimensional NumPy array representing the underlying data of the DataFrame.

  ```
  print(df.values)
  ```

  Output:

  ```
  [['Laptop' 1000 50 4.5]
  ['Tablet' 500 150 4.0]
  ['Smartphone' 800 200 4.7]
  ['Monitor' 300 100 4.3]]
  ```


- **head(n):** This method returns the first n rows of the DataFrame. If n is not specified, it defaults to 5.

```
print(df.head(2))
```

Output:

```
    Product   Price    Stock   Rating
0    Laptop    1000       50      4.5
1    Tablet     500      150      4.0
```

- **tail(n):** This method returns the last n rows of the DataFrame. If n is not specified, it defaults to 5.

```
print(df.tail(2))
```

Output:

```
       Product  Price  Stock  Rating
2  Smartphone    800    200     4.7
3     Monitor    300    100     4.3
```

Importing a CSV file into a DataFrame

This function is versatile and handles various configurations of CSV files, making it straightforward to load data for analysis or manipulation.

You can import a CSV file into a pandas DataFrame using the `read_csv()` function. The syntax of the `read_csv()` function is:

```
pd.read_csv("filename.csv")
```

where, `filename.csv` is the name of the file with `.csv` extension that you want to import.

The `pd.read_csv()` function in pandas is quite versatile and allows you to specify various parameters to customise how the CSV file is read. Some commonly used parameters are as follows:

- **filepath:** The path to the CSV file.
- **sep:** The delimiter to use for separating columns. These delimiters can be a comma, semicolon, tab, or any other character. The default value for 'sep' is a comma.
- **header:** It specifies which row to use as column names. 'header=0' means the column names are taken from the first line of the file. By default, 'header=0'.
- **index_col:** It specifies which column to use as the row labels.
- **dtype:** A dictionary where keys are column names and values are data types.
- **encoding:** It specifies the encoding to use for reading the file.
- **compression:** It specifies the compression mode for reading compressed files.

Program 55: To import a CSV file into DataFrame

```
# importing pandas library
import pandas as pd

# making data frame
#specify path of file in case file is in a different folder
df = pd.read_csv("Customer.csv",sep=',', header=0)
df.head(10)
```

Output:

```
   Customer ID First Name  Last Name           City
0            1       Akash      Patel         Mumbai
1            2       Priya     Sharma          Delhi
2            3       Aarav      Singh         Jaipur
3            4        Neha    Trivedi      Ahmedabad
4            5       Rahul      Mehta      Bangalore
5            6   Aishwarya       Iyer        Chennai
6            7      Rajesh      Reddy      Hyderabad
7            8       Sneha      Gupta        Kolkata
8            9      Sanjay      Kumar        Lucknow
9           10     Deepika      Menon           Pune
```

Exporting a DataFrame to a CSV file

The 'to_csv()' function in Pandas converts a DataFrame into CSV format. This function allows you to save the contents of a DataFrame into a CSV file with various configuration options. You can provide a file object to write the CSV data to a file; otherwise, the CSV data is returned as a string.

Program 56: To export a DataFrame to a CSV file

```
# Import pandas library
import pandas as pd

# Define a dictionary containing employee data
data = {'Name':['Adit', 'Ekam', 'Sakshi', 'Anu'],
        'Age':[27, 24, 25, 30],
        'Address':['Delhi', 'Kanpur', 'Meerut', 'Indore'],
        'Qualification':['Msc', 'MA', 'MCA', 'PhD']}

# Convert the dictionary into DataFrame
df = pd.DataFrame(data)

# Export the DataFrame to a CSV file
df.to_csv('employee_data.csv', sep=',', index=False)
```

This code creates a DataFrame from a dictionary containing employee data. It then exports this DataFrame to a CSV file named 'employee_data.csv'. The parameter 'index=False' is used to exclude the index from being written to the CSV file. The resulting CSV file will display the data in a tabular format when opened in software like Excel.

	A	B	C	D	E
1	Name	Age	Address	Qualification	
2	Adit	27	Delhi	Msc	
3	Ekam	24	Kanpur	MA	
4	Sakshi	25	Meerut	MCA	
5	Anu	30	Indore	PhD	

Digital Literacy

Video Session

Refresh your knowledge by watching the following videos:

NumPy for Beginners in 15 minutes | Python Crash Course -
https://www.youtube.com/watch?v=uRsE5WGiKWo

Pandas for Data Science in 20 Minutes | Python Crash Course -
https://www.youtube.com/watch?v=tRKeLrwfUgU

Introduction to Scikit-learn

Scikit-learn (Sklearn) is a powerful machine learning library in Python that provides simple and efficient tools for data mining and data analysis. It simplifies the process of implementing machine learning algorithms and conducting data analysis tasks in Python. Scikit-learn heavily depends on NumPy, SciPy, and Matplotlib.

Some features of the scikit-learn are as follows:

- **Simple and efficient tools:** Scikit-learn offers a simple and consistent interface for various machine learning tasks, making it easy to use and learn. It's built on top of other scientific libraries in Python such as NumPy, SciPy, and matplotlib.
- **Consistent Interface:** Scikit-learn provides a consistent API across different algorithms, making it easy to switch between different models.
- **Wide range of algorithms:** It provides implementations of various supervised and unsupervised learning algorithms, including classification, regression, clustering, dimensionality reduction, and model selection.
- **Model evaluation and validation:** Scikit-learn offers tools for model evaluation and validation, including methods for cross-validation and metrics for evaluating model performance such as accuracy, precision, F1-score, etc.
- **Data preprocessing:** It includes a wide range of preprocessing techniques for handling missing values, feature scaling, encoding categorical variables, and feature extraction.
- **Feature selection**: Scikit-learn provides utilities for feature selection and dimensionality reduction, including methods like PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), and feature importance ranking.
- **Integration with other libraries:** Scikit-learn seamlessly integrates with other Python libraries such as pandas for data manipulation, matplotlib and seaborn for data visualisation, and TensorFlow or PyTorch for deep learning.
- **Interoperability:** Scikit-learn is designed to work well with other scientific and data analysis libraries in Python, facilitating interoperability and allowing users to combine different tools seamlessly in their workflows.

You can install Scilit-Learn using pip. For installing Scilit-Learn, you need to open your terminal or command prompt and run the following command:

```
pip install scikit-learn
```

The 'Iris' Dataset

The Iris dataset is a classic dataset in machine learning and statistics. It is often used as a beginner's dataset for learning classification algorithms and data visualisation methods. The dataset consists of 150 samples of iris flowers, each with four features: sepal length, sepal width, petal length, and petal width. These samples belong to three species of iris: Setosa, Versicolor, and Virginica. Each species has 50 samples.

The goal of using this dataset is typically to develop and train classification models that can accurately predict the species of an iris flower based on its measurements. The dataset is often split into training and testing sets, with a portion of the data reserved for training the model and the remaining portion used to evaluate the model's performance.

The Iris dataset has become a standard benchmark for testing new machine learning algorithms, especially in the field of pattern recognition and classification. It's often used to demonstrate techniques such as k-nearest neighbours, decision trees, support vector machines, and neural networks.


Iris Setosa


Iris Versicolor


Iris Virginica

Brainy Fact

The Iris dataset, introduced by Ronald Fisher in 1936, is a fundamental resource in machine learning and data science. It's often considered the "Hello, World!" of machine learning due to its long-standing use in teaching statistical techniques and algorithms. With its simplicity and clear structure, it's a great starting point for newcomers to understand key data science concepts.

The Iris dataset is ordered by species. The structure of the dataset is as follows:

- **Samples 0-49:** Iris-Setosa (label 0)
- **Samples 50-99:** Iris-Versicolor (label 1)
- **Samples 100-149:** Iris-Virginica (label 2)

Let us understand how to work with this dataset.

Loading the Iris Dataset

The Iris dataset is included as one of the default datasets in scikit-learn. You can load it using the `load_iris()` function from the sklearn.datasets module. This dataset is readily available for use without needing to download or import it separately.

Program 57: To load the IRIS dataset

```
# load dataset
from sklearn.datasets import load_iris
iris = load_iris()
print(iris.data[:10])  # print the first 10 lines of the dataset
```

Output:

```
[[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.  3.6 1.4 0.2]
 [5.4 3.9 1.7 0.4]
```


```
[4.6 3.4 1.4 0.3]
[5.  3.4 1.5 0.2]
[4.4 2.9 1.4 0.2]
[4.9 3.1 1.5 0.1]]
```

In the above code, we first load the Iris dataset using the `load_iris()` function and store the dataset in the iris variable. Then, the code prints the first 10 lines of the dataset. Here, each row represents an iris flower and each column represents a feature (or a measurement of the flower). The format of the data is sepal length, sepal width, petal length, petal width (in cm). `iris.data[:10]` uses Python slicing to select the first 10 rows of the feature array. The syntax [:10] means "from the start up to, but not including, the 10th index", which effectively gives the first 10 rows (0-9 rows).

Separating Dataset into Feature and Target Value

Splitting data into training and testing sets is a critical step in machine learning to evaluate the performance of your model on unseen data. Typically, you use a portion of your data to train the model and the rest to test its performance.

In the context of supervised learning, it is common practice to separate the dataset into features and target values. The description of these values are as follows:

- **Features values:** Features are the variables or attributes that describe the characteristics of the data samples. For example, in the Iris dataset, the features are the measurements of sepal length, sepal width, petal length, and petal width. These features are used as inputs to the machine learning model to make predictions or classifications.
- **Target values:** The target values, also known as labels or classes, are the values we want the model to predict or classify. For example, in the Iris dataset, the target values represent the species of each iris flower: Setosa, Versicolor, or Virginica. The model learns to associate patterns in the features with the corresponding target values during the training process.

Program 58: To separate the data into features and target values for IRIS dataset.

```
# load dataset
from sklearn.datasets import load_iris
iris = load_iris()

# separate the data into features and target
X = iris.data
y = iris.target
```

Let us understand the above code:

X = iris.data

- This statement assigns the feature data of the Iris dataset to the variable X.
- iris.data contains the measurements for the features (sepal length, sepal width, petal length, and petal width) of the Iris flowers.
- X will be a NumPy array where each row corresponds to a sample (flower) and each column corresponds to a feature.

Y = iris.target

- This statement assigns the target labels of the Iris dataset to the variable y, i.e. Y will hold the labels that the machine will learn to predict.
- iris.target contains the labels for the species of the Iris flowers.
- Y will be a NumPy array where each element is the species label for the corresponding row in X.

Let us print the first 10 values of X and Y:

Program 59: To print the first 10 values of the separated data into X and Y in the IRIS dataset.

```
# load dataset
from sklearn.datasets import load_iris
iris = load_iris()

# separate the data into features and target
X = iris.data
y = iris.target

# print the first 10 lines of the dataset
print("Features (X):")
print(X[:10])

print("\nTarget (y):")
print(y[:10])
```

Output:

```
Features (X):
[[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.  3.6 1.4 0.2]
 [5.4 3.9 1.7 0.4]
 [4.6 3.4 1.4 0.3]
 [5.  3.4 1.5 0.2]
 [4.4 2.9 1.4 0.2]
 [4.9 3.1 1.5 0.1]]
Target (y):
[0 0 0 0 0 0 0 0 0 0]
```

As you know that, the first 50 samples in the dataset belong to the Iris-setosa species, which is why the first 10 target values are all 0.

Splitting Data For Training and Testing Set

Splitting data into training and testing sets is a critical step in machine learning to evaluate the performance of your model on unseen data. Typically, you use a portion of your data to train the model and the rest to test its performance.

The train_test_split function is a machine learning tool that divides a dataset into two parts: one for training the model and another for testing the model. The description for training data and testing data is as follows:

- **Training data (X_train, y_train):** This portion of the dataset is utilised to train the model. This data is used to train the model to identify patterns and relationships.
- **Testing data (X_test, y_test):** This portion of the dataset is used to evaluate how well the model learns from the training data. The model is evaluated on this data to determine how well it can predict outcomes for new, previously unseen data.

Splitting data into training and testing sets is essential in machine learning for several reasons:

- **Evaluation of model performance:** By splitting the dataset, you can evaluate how well your model generalises to new, unseen data. The testing set serves as a proxy for real-world data, allowing you to assess the model's performance accurately.
- **Avoiding overfitting:** Overfitting occurs when a model learns to memorise the training data's patterns instead of learning the underlying relationships. Splitting the data ensures that you can evaluate the model's performance on data it has not seen during training. If the model performs well on the testing set, it indicates that it has learned to generalise rather than memorise.
- **Model selection:** When comparing different models or algorithms, it's crucial to have a standardised testing set for fair comparison. Splitting the data ensures that each model is evaluated on the same set of unseen examples, allowing you to make informed decisions about which model performs best.

Let us now split the data into training set and testing set:

Program 60: To split the data of the IRIS dataset into training set and testing set.

```
from sklearn.model_selection import train_test_split
# load dataset
from sklearn.datasets import load_iris
iris = load_iris()

# separate the data into features and target
X = iris.data
y = iris.target

# Split the data: 80% for training, 20% for testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_
state=1)

# Splitting the data into training and testing sets (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_
state=42)

# Printing the shapes of the training and testing sets to verify the split
print("Training set - Features:", X_train.shape, " Labels:", y_train.shape)
print("Testing set - Features:", X_test.shape, " Labels:", y_test.shape)
```

Output:

```
Training set - Features: (120, 4)  Labels: (120,)
Testing set - Features: (30, 4)  Labels: (30,)
```

Adding a Classifier: KNeighborsClassifier

Scikit-learn provides a diverse set of machine learning (ML) methods, each following a standard interface for tasks such as model fitting, prediction, and performance metrics like accuracy and recall. One of the most important and commonly used method is K-Nearest Neighbors (KNN) classifier.

K-Nearest Neighbors (KNN) is a simple classification algorithm used in supervised learning. It's primarily employed for classification tasks, although it can also be adapted for regression. The goal of this classifier is to assign labels to new occurrences based on their resemblance to those in the training set.

KNN produces predictions based on the majority class of the 'k' nearest data points. This method is very useful for small to medium-sized datasets and is simple to apply and analyse.

Before using the KNN classifier, you need to first import the KNeighborsClassifier by using the following code:

```
from sklearn.neighbors import KNeighborsClassifier
```

Let us now add a KNN classifier.

Program 61: To add a KNN classifier.

```
# Import necessary libraries
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier

# Load the iris dataset
iris = load_iris()
X = iris.data  # Features
y = iris.target  # Target labels

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_
state=1)

# Create a KNN classifier with 3 neighbors
knn = KNeighborsClassifier(n_neighbors=3)

# Train the KNN classifier on the training data
knn.fit(X_train, y_train)

# Use the trained classifier to make predictions on the test data
y_pred = knn.predict(X_test)
```

Evaluating Metrics

When working with machine learning models, evaluating their performance using appropriate metrics is crucial to understand how well the model is performing and to make informed decisions about its effectiveness. Metrics evaluate how well the model makes predictions, allowing us to better understand its usefulness and identify areas for improvement.

Some important uses of metrics are as follows:

- **Model evaluation:** Metrics assist in determining how well a model works on a specific dataset. Accuracy, precision, recall, F1-score, and AUC-ROC are some of the most commonly used evaluation metrics.
- **Comparison:** Metrics enable the comparison of many models or algorithms to identify which one performs best for a given task.
- **Validation:** During model construction, metrics are used to assess the model's performance on distinct training and test sets to ensure that it generalises well to new data.
- **Optimisation:** Metrics assist hyperparameter tuning and feature selection to optimise model performance.

There are various ways by which you can evaluate the metrics. Some commonly used ways to evaluate metrics are as follows:

- **Accuracy:** This metric measures the proportion of correctly classified instances out of the total instances. In general, an accuracy of 1.0 (100%) indicates perfect classification, meaning that all instances were classified correctly. Conversely, an accuracy of 0.0 (0%) indicates that none of the instances were classified correctly.

$$\text{Accuracy} = \frac{\text{Number of correctly classified instances}}{\text{Total number of instances}}$$

- **Confusion matrix:** A confusion matrix provides a breakdown of correct and incorrect classifications for each class.

	Predicted Negative	**Predicted Positive**
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

- **Precision:** Precision measures the accuracy of positive predictions. It is the ratio of correctly predicted positive observations to the total predicted positives.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision matrix would be most appropriate to use when the cost of false positives is high.

Before evaluating the metrics, you need to first import it by using the following code:

```
from sklearn import metrics
```

In Python, the metrics.precision_score() method is used to calculate the precision of a classification model in Python using the sklearn library.

Let us now evaluate the metrics.

Program 62: To evaluate the metrics of the IRIS dataset after KNN classification.

```
from sklearn.model_selection import train_test_split
from sklearn.datasets import load_iris
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics
# load dataset
iris = load_iris()
# separate the data into features and target
X = iris.data
y = iris.target
# Split the data: 80% for training, 20% for testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_
state=1)
# Splitting the data into training and testing sets (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_
state=1)
# Create a KNN classifier with 3 neighbors
knn = KNeighborsClassifier(n_neighbors=3)
# Train the KNN classifier on the training data
knn.fit(X_train, y_train)
# Use the trained classifier to make predictions on the test data
y_pred = knn.predict(X_test)
```

```
# Calculate accuracy
accuracy = metrics.accuracy_score(y_test, y_pred)
# Calculate precision
precision = metrics.precision_score(y_test, y_pred, average='weighted')
print("Accuracy:", accuracy)
print("Precision:", precision)
```

Output:

```
Accuracy: 1.0
Precision: 1.0
```

Now, if you want to validate the predictive accuracy of the model based on the sample data. For example:

[[3, 5, 4, 2], [2, 2, 5, 4]]

Program 63: To evaluate the metrics of the IRIS dataset based on the sample data for testing.

```
from sklearn.model_selection import train_test_split
from sklearn.datasets import load_iris
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics
# load dataset
iris = load_iris()
# separate the data into features and target
X = iris.data
y = iris.target
# Split the data: 80% for training, 20% for testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
# Splitting the data into training and testing sets (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_
state=1)
# Create a KNN classifier with 3 neighbors
knn = KNeighborsClassifier(n_neighbors=3)
# Train the KNN classifier on the training data
knn.fit(X_train, y_train)
# Use the trained classifier to make predictions on the test data
y_pred = knn.predict(X_test)
# Calculate accuracy
accuracy = metrics.accuracy_score(y_test, y_pred)
# Make a prediction based on the new sample data
sample = [[3, 5, 4, 2], [2, 2, 5, 4]]
prediction = knn.predict(sample)
prediction_species = [iris.target_names[p] for p in prediction]
print("Accuracy:", accuracy)
print("Predictions:", prediction_species)
```


Output:

```
Accuracy: 1.00
Predictions: ['versicolor', 'virginica']
```

All code on the Iris dataset can be accessed through the link:

https://colab.research.google.com/drive/1IgB45-kTDe3onvY0_kDel7XwWD1RBgVb?usp=sharing

IBM Skill Build: Python for Data Science

IBM Skills Build is a comprehensive learning platform offered by IBM to help individuals develop the skills needed for various roles in technology, including Data Science, cloud computing, artificial intelligence, cybersecurity, and more. The program provides a range of courses, learning paths, and hands-on projects designed to build proficiency in specific areas of technology. Python is widely used in data science due to its simplicity, versatility, and a rich ecosystem of libraries specifically designed for data manipulation, analysis, and visualisation.

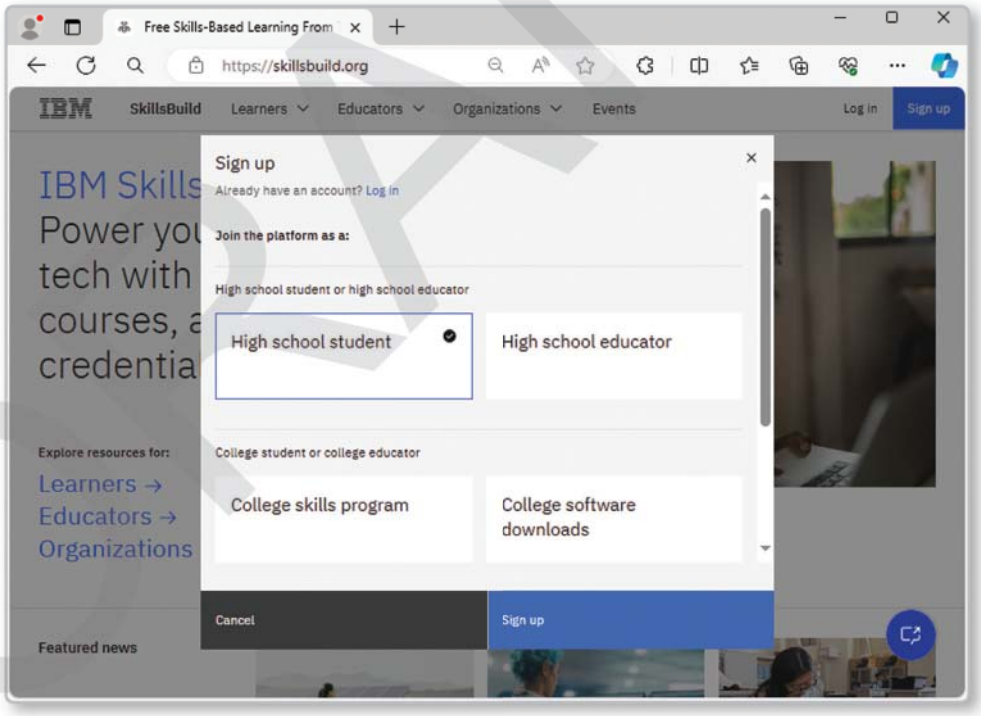
Python for Data Science is a specialised program within IBM's learning platform, designed to equip individuals with the requisite skills for proficiently leveraging Python in Data Science applications. Through a curated curriculum, hands-on exercises, and real-world projects, participants gain mastery in utilising Python's capabilities for data manipulation, analysis, visualisation, and machine learning, thereby empowering them to excel in data-driven roles.

The step to use the IBM Skills Build program to learn Python for Data Science are as follows:

Step 1: Visit the IBM Skills Build website by using the https://skillsbuild.org/ web address.

Step 2: Click on the **Sign up** button in the Home page of the IBM Skills Build website

Step 3: Select the **High school student** option to join the program as a student and click on the **Sign up** button.

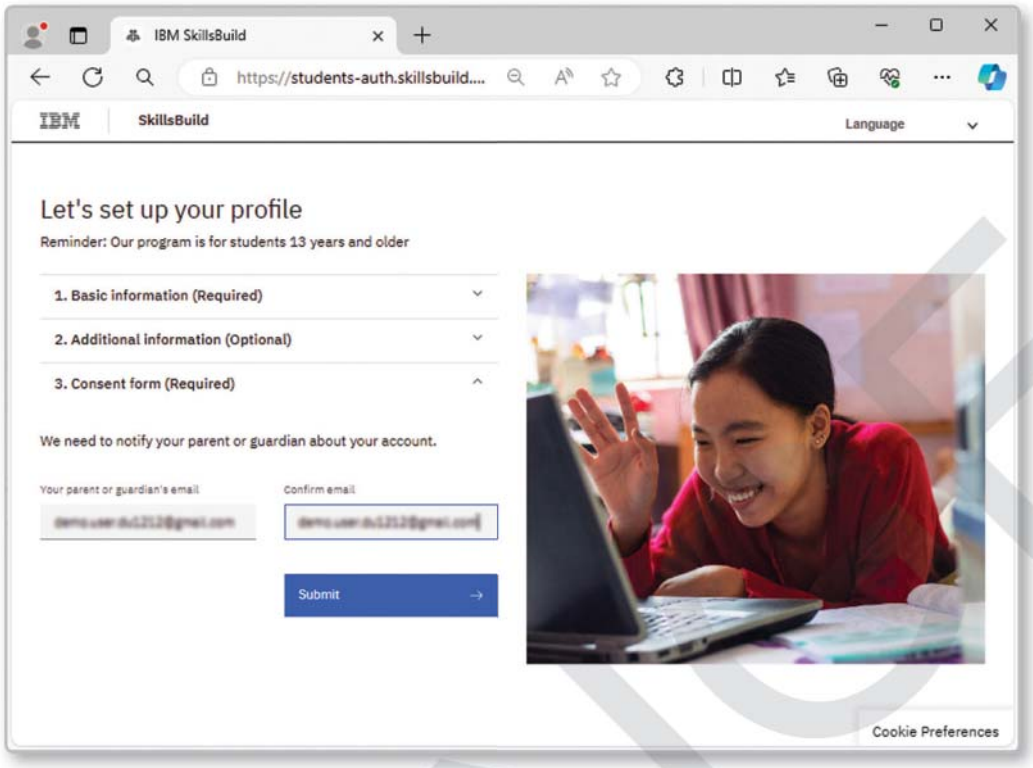


Step 4: Select the desired option from Google, LinkedIn, or IBM ID to specify how would you sign up in the IBM Skill Build website.

Or

Select the Sign up with Email option to create a new account.

Step 5: Specify the details, such as name, father name, your email id and parent email id, etc. to build your profile.

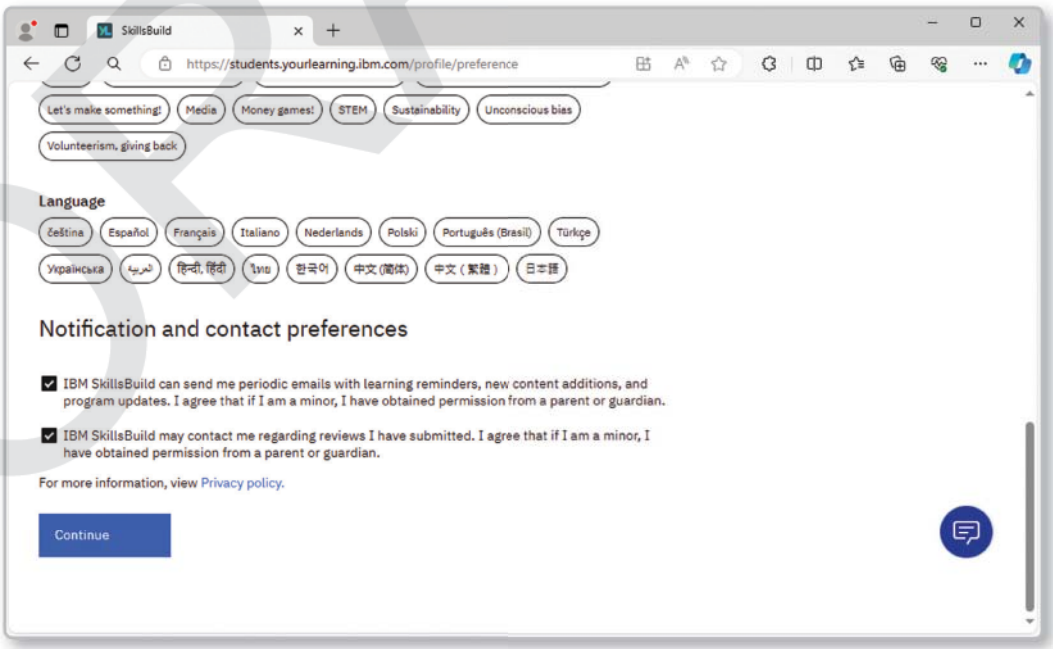


A message appears that state the confirmation mail is sent to the email of the parent for the permission to use the IBM Skill Build learning platform.

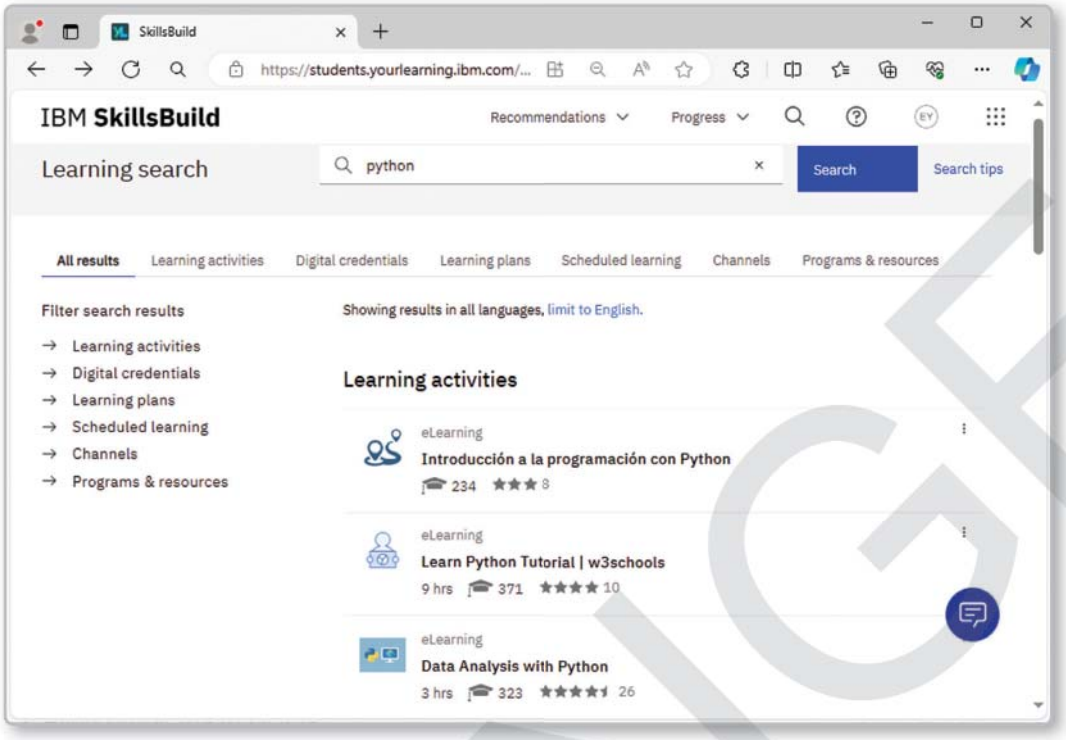
Step 6: As the parent gives the permission, an email is sent to the student email id to confirm their account and access the IBM Skill Build learning platform.

Step 7: Specify the preferred course and language in which you want to start learning.

Step 8: Accept all necessary term and condition, click on the **Continue** button to start learning.

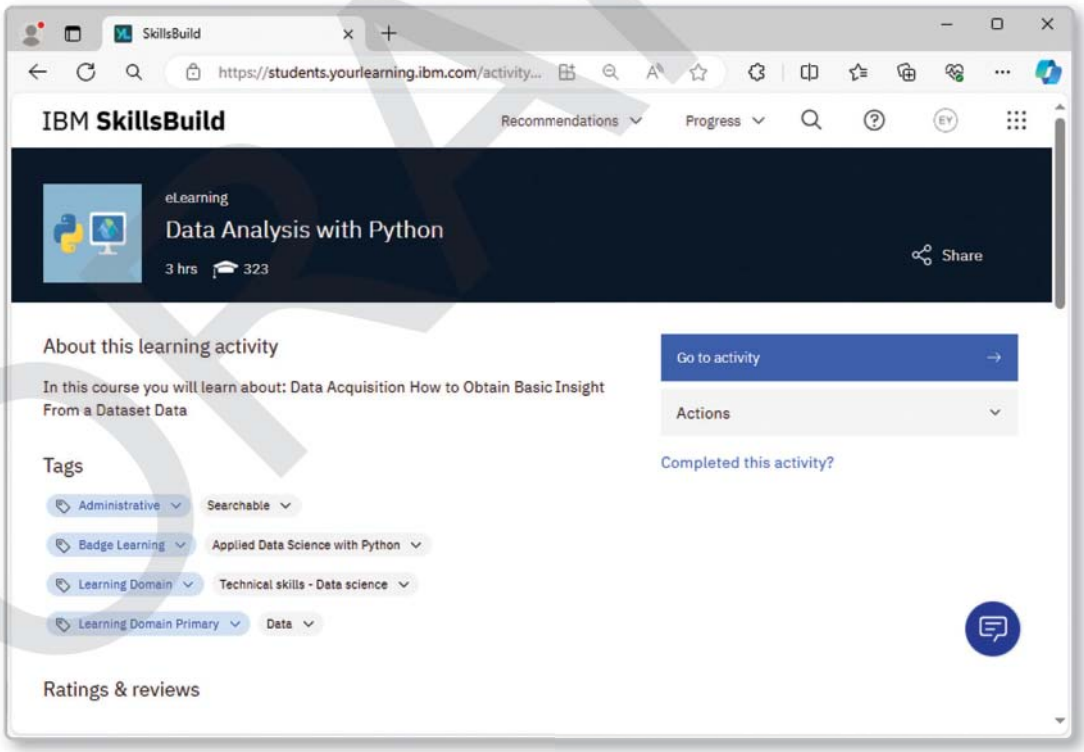


Step 9: Type python in the search text box at the top of the webpage to find relevant courses and click on the **Search** button. The list of different Python program displayed.



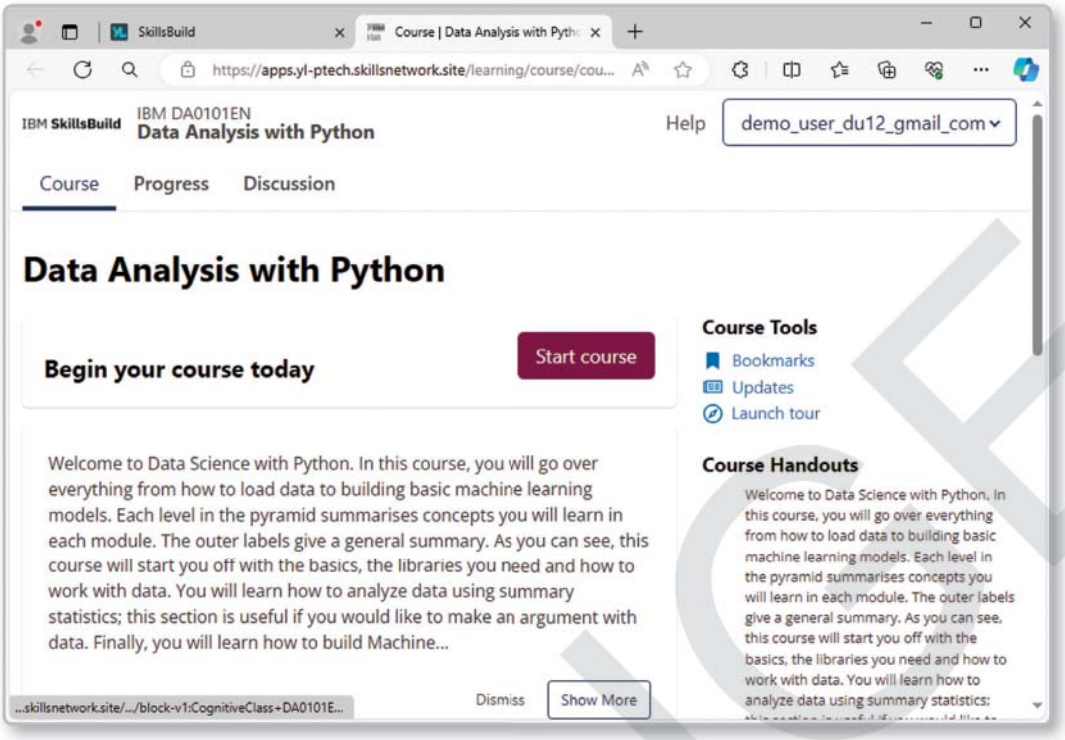
Step 10: Select the desired course that you want to learn.

Step 11: Click on the **Go to activity** button to move to the activity area of the selected program.

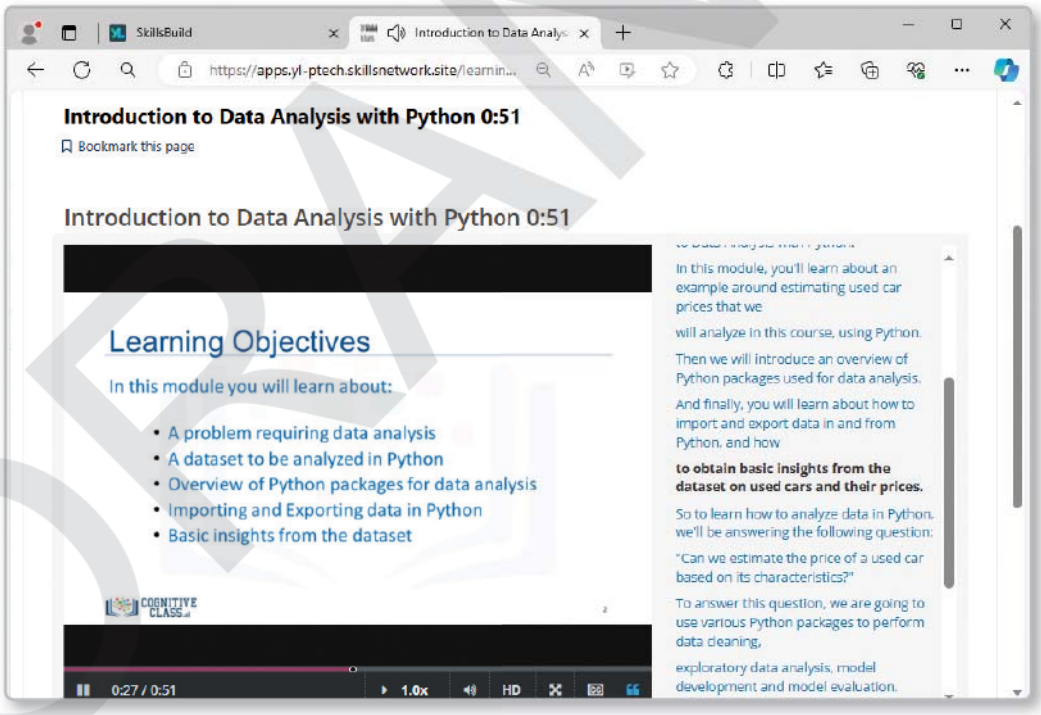


The activity page opens.

Step 12: Click on the **Start course** button to start the course.



After fulfilling the necessary option, the video related to the course begins.



Step 13: Complete the course and exercise.

Step 14: Finally, monitor the progress on the IBM Skills Build platform.

You can also explore additional courses or resources to further enhance your understanding of Python and other related topics.

At a Glance

- Python is a versatile and dynamic programming language renowned for its simplicity, readability, and extensive range of applications.
- Python has two basic programming modes—Interactive mode and Script mode.
- Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualisations, and narrative text.
- In Python, a character set refers to a collection of characters, typically defined by a specific encoding scheme, such as ASCII (American Standard Code for Information Interchange) or Unicode.
- Identifiers in Python are names given to various program elements such as variables, functions, classes, modules, etc.
- Tokens in Python are fundamental elements of the language's syntax and are the smallest individual units of a program.
- In Python, operators can be defined as special symbols which perform arithmetic and logical computation.
- In Python, comments are used to annotate code with explanations, documentation, or notes.
- Variables in Python are declared simply by assigning a value to them using the assignment operator (=).
- Data types in Python specify the type of values that a variable can hold.
- A CSV file is a popular and widely used format for storing and exchanging tabular data.
- NumPy is a fundamental library in Python that is used for performing numerical computation.
- Pandas is a popular Python library widely used for data manipulation and analysis.
- Scikit-learn is a powerful machine learning library in Python that provides simple and efficient tools for data mining and data analysis.
- The Iris dataset is a classic dataset in machine learning and statistics.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AiQuiz

A. Tick (✓) the correct option.

1. What is meant by "Python is dynamically typed language?
   a. You must declare the type of a variable when you create it. ◯
   b. The type of a variable is inferred at runtime. ◯
   c. Python does not support variable types. ◯
   d. Variable types are checked at compile time. ◯

2. Which programming paradigm does Python support?
   a. Class-oriented ◯ b. Variable-oriented ◯
   c. Object-oriented ◯ d. All of these ◯

3. Which of the following is NOT a keyword in Python?
   a. var ◯ b. async ◯
   c. def ◯ d. lambda ◯

4. Which punctuation mark is used in Python for accessing attributes and methods of objects?

   a. Comma (,) ◯ b. Period (.) ◯

   c. Colon (:) ◯ d. Semicolon (;) ◯

5. Which punctuation mark is used to separate statements on the same line in Python?

   a. Semicolon (;) ◯ b. Period (.) ◯

   c. Newline ◯ d. Comma (,) ◯

6. How many times will the while loop execute in the following program?

```
number = int(input("Enter a number to print its multiplication table: "))
counter = 1
print("Multiplication table of", number, ":")
while counter <= 10:
    print(number, "x", counter, "=", number * counter)
    counter += 1
```

   a. 9 ◯ b. 10 ◯

   c. 11 ◯ d. 0 ◯

7. How do you write multiple rows to a CSV file in Python?

   a. writer.writerow(data) ◯ b. writer.write(data) ◯

   c. writer.writerows(data) ◯ d. writer.writemultiple(data) ◯

8. What would be the output of the following code?

```
import numpy as np
arr_2d = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])
print(arr_2d)
```

   a. [1 2 3] [4 5 6] [7 8 9] ◯ b. (1 2 3) (4 5 6) (7 8 9) ◯

   c. [[1 2 3] [4 5 6] [7 8 9]] ◯ d. [[1, 2, 3], [4, 5, 6], [7, 8, 9]] ◯

9. What is the main purpose of the Pandas library?

   a. Web development ◯ b. Data manipulation and analysis ◯

   c. Game development ◯ d. Graphics design ◯

10. What are the two main data structures provided by Pandas?

    a. List and Tuple ◯ b. Array and Matrix ◯

    c. Series and DataFrame ◯ d. Dictionary and Set ◯

11. How do you select the second row (index 1) from a DataFrame using the .iloc[ ] method?

    a. df.loc[1] ◯ b. df.iloc[1] ◯

    c. df.iloc[2] ◯ d. df.select[1] ◯

12. How can you add a new column to the end of a DataFrame in pandas?

    a. Using the `.append()` method ◯

    b. Using the `.insert()` method ◯

    c. Directly assigning values to a new column using bracket notation ◯

    d. Using the `.concat()` method ◯

13. In the given example, what is the correct output for creating a DataFrame using lists or arrays?

```
import pandas as pd
data = [['Mayur', 25],
        ['Esha', 30],
        ['Vedant', 28]]
df = pd.DataFrame(data, columns=['Name', 'Age'])
print(df)
```

a.
```
     Name  Age
0  Mayur   25
1   Esha   30
2 Vedant   28
```

b.
```
     Name      Age
0    25     Mayur
1    30      Esha
2    28    Vedant
```

c.
```
     Name   ID
0  Mayur   25
1   Esha   30
2 Vedant   28
```

d.
```
ID     Name   Age
 0    Mayur    25
 1     Esha    30
 2   Vedant    28
```

14. Which method allows you to delete rows from a DataFrame?

a. `.remove()`
b. `.drop()`
c. `.delete()`
d. `.pop()`

15. Which Pandas method can be used to fill missing values with a specific value, such as the mean of a column?

a. `dropna()`
b. `fillna()`
c. `interpolate()`
d. `insert()`

16. What is the default delimiter used by the read_csv() function in Pandas?

a. Semicolon (;)
b. Tab (\t)
c. Comma (,)
d. Space ( )

17. Which of the following is a characteristic of the Iris dataset?

a. It contains data for 200 samples of iris flowers.
b. Each sample has six features.
c. The samples belong to three species of iris flowers.
d. It is primarily used for regression tasks.

18. Which of the following metrics measures the proportion of correctly classified instances out of the total instances?

a. Precision
b. Accuracy
c. Recall
d. F1-score

19. Which code snippet would you use to calculate the precision of a classification model in Python using the sklearn library?

a. `metrics.accuracy_score(y_test, y_pred)`
b. `metrics.confusion_matrix(y_test, y_pred)`
c. `metrics.precision_score(y_test, y_pred, average='weighted')`
d. `metrics.f1_score(y_test, y_pred)`

20. Which of the following metrics would be most appropriate to use when the cost of false positives is high?

a. Accuracy
b. Precision
c. Recall
d. F1-score

B. Fill in the blanks.

1. Python comes with a large standard library that provides a wide range of _____ and _____ for tasks such as file I/O, networking, web development, and more, reducing the need to write code from scratch.
2. _____ allows you to enter Python commands directly into the command line shell and immediately see the results.
3. Script mode involves writing Python code in a text file with a _____ extension and then executing that file using the Python interpreter.
4. _____ in Python are fundamental elements of the language's syntax and are the smallest individual units of a program.
5. A _____ data type refers to an ordered collection of elements, where each element is indexed by a non-negative integer.
6. The _____ in Python generates a sequence of numbers. It is commonly used with for loops to iterate over a specific range of numbers.
7. _____ involves replacing missing values with a specific value.
8. Attributes do not perform any computation or modification on the data; instead, they provide access to _____ or properties of the object.
9. Scikit-learn provides utilities for feature selection and dimensionality reduction, including methods like _____, _____ and feature importance ranking.
10. _____ produces predictions based on the majority class of the 'k' nearest data points.

C. State whether these statements are true or false.

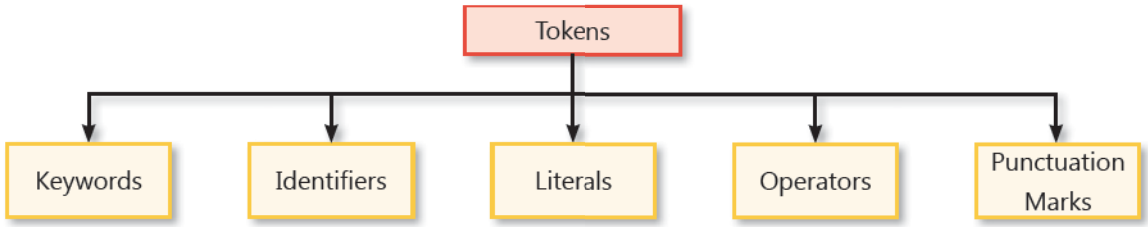
1. Python is a compiled language. _____
2. Python code is first divided into tokens during the lexical analysis phase of the interpretation process. _____
3. The values which the operators use to get the output are called identifiers. _____
4. Variables in Python do not have fixed locations. _____
5. == is an assignment operator. _____
6. The Boolean data type in Python represents one of three values: True or False or None. _____
7. Multiple assignment of variables is possible in Python. _____
8. Python libraries are designed to be reused in different projects that makes the development easier by providing various ready-to-use solutions for particular tasks. _____
9. Series represents a two-dimensional array capable of holding various types of data in it. _____
10. You can add a column in a DataFrame by directly assign values to a new column using the DataFrame's bracket notation. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions:

1. Name the types of tokens in Python.

Ans.



2. What is the difference between comparison operator and logical operators? Explain any two of each.

Ans. Comparison operators are used to compare values of the operand. They return True if the comparison is true, otherwise False. They are commonly used in conditional statements and loops.

Example:

==	Checks if two operands are equal
!=	Checks if two operands are not equal

Logical operators are used to combine conditional statements. They return True or False depending on the conditions. Logical operators include AND, OR, and NOT.

Example:

and	Returns True if both operands are true
or	Returns True if at least one operand is true

3. What are the naming conventions for identifiers in Python?

Ans. The following are the identifier naming conventions:

- Must begin with a letter (a-z, A-Z) or an underscore (_).
- Subsequent characters can be letters, digits (0-9), or underscores.
- Case-sensitive (myVar is different from myvar).
- Cannot be a Python keyword (reserved words).
- No special characters such as !, @, #, $, %, etc., are allowed within identifiers.
- Blank spaces within an identifier are disallowed

5. What is a CSV file?

Ans. A CSV (Comma-Separated Values) file is a popular and widely used format for storing and exchanging tabular data. In its simplest form, each line of a CSV file represents a single record, with individual data fields separated by commas.

6. What is the syntax of importing a CSV file?

Ans. `pd.read_csv("filename.csv")`

B. Long answer type questions:

1. What are the features of Python?

Ans. Some of the important features of Python are as follows:

- **Easy to read and write:** Python's syntax is designed to be simple and easy to understand, making it accessible to beginners and experienced programmers. Its code is often referred to as "executable pseudocode."
- **Interpreted language:** Python is an interpreted language, meaning that code is executed line by line, which allows for easier debugging and prototyping.
- **Dynamic typing:** Python is dynamically typed, meaning you don't need to declare the type of a variable when you create one. The type of variable is inferred at runtime.
- **Extensive standard library:** Python comes with a large standard library that provides a wide range of modules and packages for tasks such as file I/O, networking, web development, and more, reducing the need to write code from scratch.
- **Open source:** Python is open source, meaning that its source code is freely available and can be modified and redistributed. This fosters collaboration and innovation within the Python community.
- **Object-oriented:** Python supports object-oriented programming (OOP) programming paradigm, allowing you to create classes and objects, encapsulate data, and implement inheritance and polymorphism.
- **Cross-platform:** Python code can run on various platforms and operating systems without modification, including Windows, macOS, Linux, and others.
- **Community support:** Python has a vast and active community of developers who contribute to libraries, frameworks, nd resources. This community support makes it easy to find solutions to problems and learn from others.

- **Easy integration:** Python effortlessly integrates with other programming languages and technologies, facilitating interoperability and compatibility with existing systems. It supports integration with C/C++, Java, .NET, and other languages that allows developers to use their preferred tools and technologies within Python-based projects.
- **Strong industry adoption:** Python is widely adopted across various industries, including technology, finance, healthcare, education, and more. Its robustness, scalability, and extensive library support make it an attractive choice for businesses seeking efficient and cost-effective solutions to their software development needs.

2. Why CSV files are famous and what are some basic operations of CSV files?

Ans. Due to its text-based nature, CSV files are platform-independent and easily readable by both humans and machines. This simplicity, combined with its compatibility with various programming languages and tools, makes CSV a go-to choice for data interchange, storage, and analysis across diverse domains such as finance, research, and software development.

Some basic operations of CSV files are as follows:

- **Importing the CSV library:** In Python, the csv module provides functionality to work with CSV files. It includes classes to read and write tabular data in CSV format.
- **Opening a CSV file in reading mode:** When opening a file in reading mode ('r'), you are telling Python that you only intend to read from the file, not modify it. The `csv.reader()` function then reads the contents of the file line by line.
- **Opening a CSV File in writing mode:** When opening a file in writing mode ('w'), you are telling Python that you intend to write to the file. If the file already exists, it will be truncated (emptied) first. If it doesn't exist, a new file will be created.
- **Closing a CSV file:** In Python, it is important to close files after you have finished working with them. However, using the with statement automatically closes the file when the block is exited, so you don't have to worry about explicitly closing it.
- **Writing rows to a CSV file:** You can use the `writerows()` method to write multiple rows to a CSV file at once. Each row should be a list of values. This function will replace all existing data in the CSV file.
- **Append rows to a CSV file:** Appending a row to an existing CSV file involves opening the file in append mode ('a'), and writing a single row of data to it. You can achieve this using the csv.writer object's `writerow()` method.

3. List and explain several ways of handling missed values in a DataFrame.

Ans. There are several ways by which you can handle missing values in DataFrame:

- **Imputation:** Imputation involves replacing missing values with a specific value. Common strategies include replacing missing values with the mean, median, or mode of the column. This method helps in retaining the structure of the dataset and avoids losing valuable information.

  Pandas provides methods like `fillna()` to perform imputation. For example, you can fill missing values in a DataFrame `df` with the mean of each column using `df.fillna(df.mean())`.

- **Dropping:** Dropping involves removing rows or columns containing missing values. This method is useful when the missing values are sparse and dropping them doesn't significantly impact the analysis.

  Pandas provides the `dropna()` function to drop rows or columns with missing values. For example, you can drop rows with any missing values in a DataFrame `df` using `df.dropna()`.

- **Interpolation:** Interpolation involves estimating missing values based on existing data. Pandas provides interpolation methods such as `interpolate()` to estimate missing values. For example, you can perform linear interpolation on a DataFrame `df` using `df.interpolate()`. This method is particularly useful for time series or ordered data where missing values can be inferred from neighbouring values.

4. Explain and exemplify multiple assignments in Python.

Ans. Multiple assignment of variables, also known parallel assignment, is a powerful feature in many programming languages that allows you to assign multiple variables at once.

For example:

x, y, z = 1, 2, 3

In given code, the values 1, 2, and 3 are assigned to variables x, y, and z, respectively. This simultaneous assignment saves lines of code and improves readability compared to assigning each variable separately.

Multiple assignment is not limited to simple variables. It works with any iterable data structure, including tuples, lists, and even custom objects that support iteration and unpacking.

For example:

point = (5, 10)

x, y = point

In given code, the values (5, 10) are unpacked from the point tuple and assigned to variables x and y respectively

5. What is the use of NumPy?

Ans. NumPy is the short form of Numerical Python. It is a fundamental library in Python that is used for performing numerical computation. It provides support for arrays, matrices, and a variety of mathematical functions to operate on these data structures efficiently. Its array-based data structures and operations execution makes it very useful for various applications, such as data analysis, machine learning, scientific computing, etc.

C. Competency-based/Application-based questions:

In ABC company the employees were being lectured about evaluation of an AI model. A student was confused between testing and evaluation. Help him solve his problem by giving concrete points and letting him know the importance of various metrices for evaluation of an AI model.

Ans. First let us clarify the student's doubt that testing is actually running the test cases (sample inputs to check if the results are accurate) on the developed model whereas evaluation is finding inference (understanding) out of what results have come, when test cases are run.

Testing is actually running the model in different scenarios whereas evaluation is preparing a report of what testing results are pointing out.

In machine learning, metrics are used to analyse and measure the performance of models. Metrics evaluate how well the model makes predictions, allowing us to better understand its usefulness and identify areas for improvement.

Some important uses of metrics are as follows:

- **Model evaluation:** Metrics assist in determining how well a model works on a specific dataset. Accuracy, precision, recall, F1-score, and AUC-ROC are some of the most commonly used evaluation metrics.
- **Comparison:** Metrics enable the comparison of many models or algorithms to identify which one performs best for a given task.
- **Validation:** During model construction, metrics are used to assess the model's performance on distinct training and test sets to ensure that it generalizes well to new data.
- **Optimisation:** Metrics assist hyperparameter tuning and feature selection to optimize model performance.

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. Python is classified as which type of language?

a. Compiled ◯ b. Interpreted ◯

c. Assembly ◯ d. Machine ◯

2. Which feature of Python significantly reduces the need to write code from scratch?

   a. Dynamic typing ◯ b. Open source ◯

   c. Extensive standard library ◯ d. Cross-platform compatibility ◯

3. Which of the following characters is NOT part of Python's character set?

   a. @ ◯ b. $ ◯

   c. ~ ◯ d. § ◯

4. What is an example of a valid identifier in Python?

   a. 1variable ◯ b. my-variable ◯

   c. calculate_sum ◯ d. my@var ◯

5. Identify the relational operator from the following:

   a. and ◯ b. + ◯

   c. >= ◯ d. ** ◯

6. What is the type of the variable x after the following code executes?

```
x = None
```

   a. int ◯ b. float ◯

   c. NoneType ◯ d. bool ◯

7. What will be the output of the following code?

```
for i in range(1, 6):
    if i == 3:
        continue
    print("Iteration:", i)
```

   a. Iteration: 1 Iteration: 2 Iteration: 3 Iteration: 4 Iteration: 5 ◯

   b. Iteration: 1 Iteration: 2 Iteration: 4 Iteration: 5 ◯

   c. Iteration: 1 Iteration: 2 ◯

   d. Iteration: 3 Iteration: 4 Iteration: 5 ◯

8. On top of which library is Pandas built?

   a. Matplotlib ◯ b. Scikit-learn ◯

   c. NumPy ◯ d. TensorFlow ◯

9. What does the following code do?

```
import csv
with open('Customer.csv', 'a', newline='') as file:
    writer = csv.writer(file)
    writer.writerow(['4', 'Amit', 'Verma', 'Pune'])
```

   a. Reads a CSV file ◯ b. Writes a new CSV file ◯

   c. Appends a row to an existing CSV file ◯ d. Deletes a row from a CSV file ◯

10. What method is used to access a specific element in a DataFrame by its row and column index?

    a. `.loc[]` ◯ b. `.iloc[]` ◯

    c. `.at[]` ◯ d. `.ix[]` ◯

11. What is the correct syntax to create a DataFrame using a list of dictionaries?

a.
```
import pandas as pd
data = [{'ID': 1, 'Name': 'Arti', 'Age': 25},
        {'ID': 2, 'Name': 'Trinabh', 'Age': 30},
        {'ID': 3, 'Name': 'Surbhi', 'Age': 28}]
df = pd.DataFrame(data)
print(df)
```

b.
```
import pandas as pd
data = [1, 'Arti', 25, 2, 'Trinabh', 30, 3, 'Surbhi', 28]
df = pd.DataFrame(data)
print(df)
```

c.
```
import pandas as pd
data = ['ID', 'Name', 'Age']
df = pd.DataFrame(data)
print(df)
```

d.
```
import pandas as pd
data = {'ID': [1, 2, 3], 'Name': ['Arti', 'Trinabh', 'Surbhi'], 'Age': [25, 30, 28]}
df = pd.DataFrame(data)
print(df)
```

12. What is the method to add a new row to a DataFrame at the end?

a. `.add_row()` b. `.append_row()`

c. `.loc[]` d. `.concat()`

13. How can you delete multiple columns with specific labels from a DataFrame?

a. Using .remove() method b. Using .drop() method with a list of column labels

c. Using .delete() method d. Using .pop() method

14. What does the shape attribute of a DataFrame return?

a. The data types of each column.

b. The total number of elements in the DataFrame.

c. A tuple representing the dimensions of the DataFrame.

d. The underlying NumPy array of the DataFrame.

15. Which of the following commands will load a CSV file into a DataFrame in Pandas?

a. `pd.to_csv('filename.csv')` b. `pd.load_csv('filename.csv')`

c. `pd.read_csv('filename.csv')` d. `pd.import_csv('filename.csv')`

16. What does an accuracy score of 1.0 indicate about a machine learning model's performance?

a. All instances were classified incorrectly

b. The model has perfect classification

c. The model has an equal number of false positives and false negatives

d. The model needs more training data

B. Fill in the blanks.

1. Python has a vast and active community of developers who contribute to _____, _____, and resources.
2. IDLE stands for _____ .
3. The _____ function takes the user's input while a program executes.
4. A _____ is the smallest unit in a Python program.
5. The _____ statement helps us to test multiple conditions and follows a top-down approach.
6. In a dictionary, keys are _____ whereas values are mutable.
7. Pandas provides interpolation methods such as _____ to estimate missing values.
8. The delimiter is used for separating columns. These delimiters can be a comma, semicolon, tab, or any other character. The default value for 'sep' is a _____ .
9. _____ statement assigns the feature data of the Iris dataset to the variable X.
10. _____ Build is a comprehensive learning platform offered by IBM to help individuals develop the skills needed for various roles in technology, including data science, cloud computing, artificial intelligence, cybersecurity, and more.

C. State whether these statements are true or false.

1. The `print()` function prints or sends the output to the standard output device, which is usually a speaker. _____
2. A literal refers to a constant or a fixed value. _____
3. Comments are ignored by the Python interpreter during execution and are solely for human readers. _____
4. Multiple assignment works with any iterable data structure, including tuples, lists, and even custom objects that support iteration and unpacking. _____
5. Data types in Python specify the values that a variable can hold. _____
6. The while statement executes a set of statements repeatedly, until the logical expression evaluates to true. _____
7. CSV files are platform-dependent and difficult to read by both humans and machines. _____
8. `np.empty()` function creates an array without initialising its values and not allocating any memory location. _____
9. You can create a DataFrame only from lists and dictionaries. _____
10. Pandas provides the `dropna()` function to drop rows or columns with missing values. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions:

1. What do you understand by operator precedence? Explain with the help of an example.
2. Differentiate between single line and multiline comment.
3. What does KNN stand for?
4. What is dynamic typing? How is it helpful?
5. Explain the purpose of keys in a dictionary.
6. What are some real-life applications where NumPy library is useful?
7. Demonstrate few ways to create NumPy Arrays.

8. What is a DataFrame? How is it created? Explain any one way.
9. What is Iris DataSet? What is the purpose of creating it?
10. What are feature and target value of a dataset?

B. Long answer type questions:

1. Discuss the features of Scikit learn library.
2. Why is splitting data into training and testing sets is essential in machine learning?
3. Why do we use precision, confusion metrics and accuracy? How do we use them?

C. Competency-based/Application-based questions:

1. Some students of class 12th engaged themselves in a conversation and compared DataFrames with Excel. They were firm that DataFrames have no advantage over Excel. Try to find if there are any points supporting DataFrames in the discussion.
2. Students of class 12th were unwilling to use Iris dataset for their model testing. How would you ask them to reconsider their choice. Give convincing arguments so that they use this dataset for their AI model.

Ai In Life

Creativity and Innovativeness

Find out why Iris dataset in important in real life. Does it have any drawbacks? Also find out 5 alternative datasets to Iris.

Ai Deep Thinking

Problem solving & Logical Reasoning

Find out the difference between NumPy and Pandas – their data compatibility, performance, memory usage, etc.

Ai Lab

Coding & Computational Thinking

Write code in Python for the following:

1. Create a NumPy array of 10 zeros.
2. Create a NumPy array with values ranging from 40 to 60.
3. Create a Pandas DataFrame from a dictionary having Name, Age and City as keys. Assume your own values.
4. Display 'Name' and 'City' columns from the DataFrame created in the previous question.
5. Add a new column 'Marks' with the values [68, 40, 78, 82] to the DataFrame.
6. Delete the 'Age' column from the DataFrame.
7. Create a simple DataFrame. Display the row labels and columns data types.

8. Use the Iris dataset to do the following:
   a. Load the Iris Dataset using Scikit-Learn.
   b. Print the first 5 rows of the dataset.
   c. Print the names of the features (attributes) in the dataset.
   d. Print the target variable names (species) in the dataset.
9. Create a DataFrame from the Iris Dataset with column names. Print the summary statistics of the dataset (mean, min, max, etc.) for each feature.
10. Predict the output:

a.
```
x = 75
if x > 100:
    y = x * 0.20
elif x > 50:
    y = x * 0.10
else:
    y = x * 0.05
print(y)
```

b.
```
count = 0
for i in range(1, 6):
    if i % 2 == 0:
        count += 1
    else:
        count += 2
print(count)
```

c.
```
a = 3
b = 4
for i in range(1, 4):
    if a > b:
        a -= 1
    elif a < b:
        b -= 1
    else:
        a += 1
print(a, b)
```


Answers

Exercise (Section A)

A. 1. b 2. c 3. a 4. b 5. a 6. b 7. c 8. c 9. b 10. c
11. b 12. c 13. a 14. b 15. b 16. c 17. c 18. b 19. c 20. b

B. 1. modules, packages 2. interactive mode 3. .py 4. Tokens
5. Sequence 6. `range()` function 7. Imputation 8. metadata
9. PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis) 10. KNN

C. 1. False 2. True 3. False 4. True 5. False 6. False
7. True 8. True 9. False 10. True

Learn to create small games using Python!

Watch the given video and create a small puzzle game using Python.

https://medium.com/@waleedmousa975/creating-a-puzzle-game-with-python-and-machine-learning-a-step-by-step-tutorial-281b92e1dc53

Follow the steps in the tutorial and build your own game!

Answer the following questions based on the Iris Dataset.

1. Find out some popular machine learning algorithms used with the Iris dataset.

2. Also find out, whether the Iris dataset is suitable for more advanced machine learning tasks.

3. What are some of the best practices for using machine learning on the Iris dataset?

UNIT-4

INTRODUCTION TO CAPSTONE PROJECT

Learning Outcomes

- Design Thinking
- Identifying the Problem to Solve
- Empathy Map
- Understanding the Capstone Project
- Defining the problem
- Right Questioning
- Ideate
- Sustainable Development Goals
- Problem Decomposition
- Sample Capstone Projects

Creative thinking means to think in order to make something innovative. Whereas, critical thinking means "thinking about thinking". It means you can understand the way your perception works in order to recognise mistakes in your reasoning and avoid biases.

In this unit, you will learn about design thinking and its different stages. You will learn about right questioning, 5W1H problem canvas and identifying the problem to solve. At the end of this unit, you will learn about Empathy Map.

Ai Task

Recall the story of 'The Thirsty Crow' and answer the following questions:

1. What was the problem in front of the crow?

2. What techniques did the crow use to solve the problem?

3. Was the problem solved in one go or the crow had to undertake several steps?

4. Was the solution simple or complex?

5. Do you feel that the crow used out-of-the-box thinking?

Design Thinking

Design thinking is a process to solve problems creatively. It solves problems by putting consumers' needs first. So, design thinking is "people-oriented", which means that it uses evidence of how consumers (people) interact with a product or service, rather than how organisations think they will handle it. The essence of design thinking is human-centric and user-specific.

To be truly people-oriented, designers observe how people use products or services, and further refine products or services to improve the customer experience. This is the "iteration" part of design thinking.

The design thinking ideology asserts that practical, user-centric problem-solving methods can bring innovation, and innovation can bring diversity and competitive advantage. This process consists of five different stages, as shown below:

Let us learn about them in detail.

Empathise

In the first stage, designers observe consumers to gain a deeper understanding of how they interact with or are affected by the product or problem. Observations must be made with **empathy**, that is, without making any judgements or conveying any preconceived notions about consumer needs. By understanding the person affected by a problem, you can find a more impactful solution. Empathy observation is powerful because it can uncover issues the consumers didn't even know they had or that they could not verbalise. From this point on, it is easier to understand human needs.

Define

In the second stage, you collect the observations from the first stage to define the problem to be solved. This stage is about narrowing down the focus of the design thinking process. Think about the difficulties your consumers may face, the problems they face again and again, and what you have learned from how they were affected by the problems. Once you have summarised your results, you can define the problem statement.

Ideate

The next step is to gather ideas to solve the problem you identified. These brainstorming sessions can be conducted in a small group, and team can meet in an office area that encourages creativity and collaboration, an innovation laboratory, or the team members can work alone. The important thing is to generate many different ideas.

The goal is to ultimately overcome cognitive fixedness and devise new and innovative ideas that solve the problems you identified. At the end of this process, you will have some ideas to move forward.

Prototype

This is the stage where an idea becomes a real solution. A prototype is a simple experimental model for a proposed solution. The prototype may not be perfect. The focus of the prototype is to quickly come up with a concrete version of the idea and see how it will be accepted by consumers. This step isn't about perfection, but rather, experimenting with different ideas and seeing which parts work and which don't. Examples of prototypes include a website to test consumer demand for products, a 3D printed product, etc.

Test

Once you have made a prototype solution available to consumers, you need to observe how they interact with it. During this testing phase, you collect feedback about your work.

The design thinking process is iterative. At the end of the fifth stage, you may have to return to one or more stages. Perhaps the test indicates that you need to develop another prototype, so, you will return to the fourth stage. Or maybe it turns out that you misunderstood the needs of consumers, then, you need to go back to the early stages of the process.

Brainy Fact

Stanford University offers a course called 'Designing Your Life', which applies design thinking to building a joyful career and life. This is one of the most popular courses!

Digital Literacy

Video Session

Scan the QR code or go to the link to watch the video and answer the questions given below: Design Thinking Animation

https://www.youtube.com/watch?v=uRtAzzitBmA

What are the five steps discussed in this video to understand the design thinking process?

Ai Reboot

1. Design thinking is a/an _____ process. (Fill in the Blank)
2. The DT process can bring about innovation, diversity, and competitive advantage. _____ (State True or False)
3. Which step of the DT process involves brainstorming?
4. What is a prototype?
5. Using Design thinking, give a solution to the problem "Creating a Better Toothbrush".

Ai Task

Netflix has become an industry leader by consistently applying design thinking. Initially, it revolutionised movie rentals by offering a DVD subscription service that eliminated the need for physical store visits. As DVDs became outdated, Netflix transitioned to on-demand streaming to maintain its edge. The company continued to innovate by producing original content and improving user experience with short trailers, all driven by a keen focus on customer needs through design thinking.

Find out two more such companies which used design thinking to become more successful.

Right Questioning

When developing solutions using the design thinking framework, designers need to interact with customers frequently to collect detailed facts about problems and user expectations. This requires asking various questions. Great questions often lead to a better understanding of the problem, so that a good solution can be found.

To gather relevant facts and information from customers, a simple and efficient method of questioning, called the **5W and 1H method**, may be used.

WHO

- Who are the users?
- What are their features/qualities?
- If the users belong to different groups, which user group is more important?

WHAT

- What do the users do?
- What are the tasks they can accomplish?
- What extra tasks can the users do?

WHERE

- Where to perform the task?
- Where can the task be performed better (location change)?
- Lighting conditions, temperature, noise, physical movement, and other conditions of the location can impact how to perform tasks.

WHEN

- When do the users perform these tasks?
- Everyday, sometimes, what is the frequency of performing tasks?
- Can time change bring about a positive effect in the task?

WHY

- Why do the users need to perform these tasks?
- Why can't they change the way they perform the task?
- What are they really trying to accomplish?

HOW

- How to perform the task? (This involves the steps to perform the task)
- Can this method be used in other areas?

Example of 5W and 1H Problem Canvas

Goal: To explain the impact of climate change due to the increase in Carbon dioxide gas.

WHO (Stakeholders)	Concerns everyone—People, animals
WHAT (issue, problem, need)	Extreme weather conditions, Health issues, major shifts in climate
WHEN (time)	all through the year, more in winters.
WHERE (location/situation)	Everywhere—Carbon dioxide emissions from burning, pollution due to increasing vehicles on the road, gases emissions from factories, deforestation, Landfill.
WHY (reason)	People facing health issues, respiratory diseases, life span has gone down, floods, droughts.
HOW (Solution will solve/ improve problem)	Health is a major concern. If this problem is addressed properly then so many problems which are dependent on this will also improve. Clean air will improve the quality and span of life.

Identifying the Problem to Solve

Problem-solving is the process of defining the problem, determining the cause of the problem, brainstorming to generate possible solutions, and selecting alternatives for the most suitable solution. Problem-solving is at the core of a person's daily work. Whether you are solving problems for customers (internal or external) or discovering new solutions to the problems you are facing, it can be large or small, simple or complex.

Identifying a problem is more important than finding a solution to a problem. The Swacch Bharat Abhiyan kickstarted because the problem of people defecating openly was identified as the root cause of many other problems.

The **'Define'** phase of design thinking (identifying the problem) ensures that you fully understand the goals of the design project. It can help you clarify your design issues and provide a clear work goal. Without a clear problem statement, it is difficult to know what your goals are.

Ideate

Idea generation is a creative process (such as brainstorming) where designers generate ideas in meetings. This is the third stage of the Design thinking process. Participants assemble with an open mind and put forward as many ideas as possible to solve problems in a convenient and judgment-free environment.

Ideation will help you:

- Ask the right questions and innovate, focus on the users, their needs, and your understanding of them.
- Combine the views and strengths of team members.
- Get rid of the obvious solutions in your mind and let the team surpass them.

Ideation Techniques

The most essential ideation techniques used to generate various ideas are brainstorm, brain dump and brainwriting. Let us learn about them in detail.

Brainstorm

It is very popular with design teams because they can expand in all directions. Although the team has rules and moderators to keep them on track, they can use out-of-the-box lateral thinking to come up with the most effective solution to any design problem. Through brainstorming, they can take multiple approaches—the more, the better—instead of just exploring traditional methods and encountering the obstacles that come with them. A large number of ideas are collected so that various options are available for solving the problem.

Brain dump

Brain dump is very similar to brainstorm, however, it is done individually. It enables the people involved to open their hearts, put down their thoughts on a piece of paper. Participants write down their thoughts on paper or post-it notes and then share their thoughts with the larger group at a later stage.

Brainwriting

Brainwriting is also very similar to a brainstorming session and is called "individual brainstorming." In this case, only the most confident team members share their thoughts, while introverts keep their thoughts private. Brainwriting gives introverts time to write them down instead of sharing their thoughts aloud with the team. Participants write their thoughts on paper, a few minutes later, they give their notes to another participant, and then the second team member jots down his ideas or adds on to the previous member's ideas, and so on. In this way, all participants pass on their records to others. After about 15 minutes, the papers are collected and are available for immediate discussion.

Empathy Map

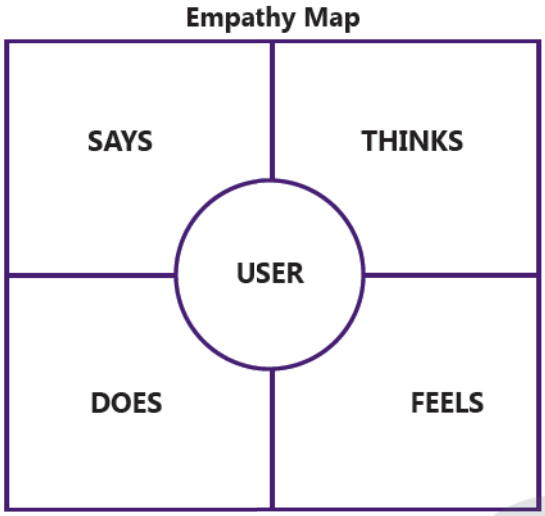
Empathy is the first step in design thinking because it enables designers to understand, empathise, and share user feelings. Through empathy, we can put ourselves in consideration for others and understand how they feel about their problems, or situations.

Before we begin to figure out what the problem is or try to solve it, it's always a good idea to "walk a mile from the user's point of view" and understand the user. A very useful tool to do this is the Empathy Map. An Empathy Map is a collaborative visualisation used to clarify our understanding of a specific type of user. It helps to:

- Create a common understanding of user needs.
- Help in decision-making.
- Deepen understanding and gain insight into user behavior.

The traditional Empathy Map is divided into 4 quadrants (Says, Thinks, Does, and Feels), with the user or person in the middle. Empathy Maps provide an overview of the user as a whole and are not sequential.

- The ***Says*** quadrant contains what the user says aloud in an interview.
- The ***Thinks*** quadrant contains what the user is thinking about his/her experience.
- The ***Does*** quadrant depicts the actions of the user.
- The ***Feels*** quadrant denotes the user's emotional state, often represented as an adjective along with a short sentence for context.



To create an Empathy Map about a user, first, draw the grid. Secondly, add the observations in the respective quadrants. These observations can be written directly in the quadrant, or one sticky note per observation can be added to each quadrant. Let's see an example, as given below.

Empathy Map for a parent, Ms. Neena, seeking admission for her ward in a school in class IX:

SAYING	THINKING
I have few queries. I need to know. I am in the process of selecting a good school for my ward. How to get information about the admission process?	How time-consuming will it be? I must get an answer to my queries quickly. Will I be able to get a seat? Will the school reception be open to answer my queries?
Ms. Neena	
DOING	**FEELS**
Applying through email for ward's admission. Calling and checking at the school admission desk. Surfing net and school websites.	I am confused. I don't understand. I am not getting enough information. I am feeling frustrated.

As the name suggests, Empathy Maps simply help us build empathy with our end users. When based on real data and when combined with other mapping methods, they can:

- Remove bias from our designs and align the team on a single, shared understanding of the user.
- Discover weaknesses in our research.
- Uncover user needs that the user themselves may not even be aware of.
- Understand what drives users' behaviours.
- Guide us towards meaningful innovation.

Digital Literacy

Video Session

Scan the QR code or visit the following link to watch the video: What is an Empathy Map?

https://www.youtube.com/watch?v=QwF9a56WFWA

After watching the video, answer the following question:

What is the purpose of an Empathy Map?

Sustainable Development Goals

The Sustainable Development Goals (SDGs) are also termed as the Global Goals. These goals were adopted by the United Nations in 2015. The goals were adopted in New York as a stand for humanity, as a universal call to take action to solve the problems of poverty, malnourishment, poor climate, and hygiene throughout the world and bring peace, equality, growth, availability of energy resources, and partnership to achieve a well balanced world.

The SDG goals are interrelated as they affect each other, and one cannot be completely achieved without incorporating the other. These goals bring balance to different spheres of life like-social, economic, and environmental front will achieve stability when we strive to work towards the SDGs.

Contributions from financial, technical, and creative resources are necessary to achieve the SDGs. The SDGs give us a direction to work towards the goals to reach for equality and prosperity we are looking for in the entire world. The countries have to collaborate for the development of all and prioritise taking those countries along that do not have appropriate resources for the SDGs and work for a healthy world overall.

Goal 1: No Poverty

Poverty is one of the greatest challenge that the world has to overcome for many decades. Various institutions are helping to reach this goal and they have seen success too. The poverty rate has dropped drastically since 1990, but overall, there are still so many people struggling with even the most basic human needs.

At present, more than 700 million people are living in extremely poor conditions, and the count is higher than in 2019. Though poverty has dropped overall, the number of people facing a dearth of basic facilities is still very high.

Extreme poverty is seen in Sub-Saharan countries, and poverty is decreasing in East Asian countries. China and India have been the most successful of all, in uplifting the people out of poverty.

Pointing out the same, the SDGs have directed themselves to eradicate poverty as much as possible, saying that there should be no one below the poverty line after 2030. The goal is to work for the vulnerable and poorest of the poor groups.

Goal 2: Zero Hunger

People are getting better nutrition all over the world because of improvements in agricultural productivity. Poor countries like Central and East Asia and Latin America that used to suffer from famines and hunger have made improvements in providing appropriate nutrition to their populations. Somehow, many countries are still facing the issue of hunger for their people.

Worldwide, more than 800 million people are undernourished, and over 90 million children are dangerously malnourished. Food insecurity has increased in African countries, and a lot needs to be done to manage it. Even South American countries are working to uplift themselves from food scarcity because their children are undernourished. Thus, the United Nations has considered it and made "zero hunger" one of its goals.

The SDGs aim to end all forms of hunger and malnutrition by 2030, making sure all people—especially children—have sufficient and nutritious food all year.

This SDG will require consistent support for the farmers by their government and the employment of all the means, like providing technological access, markets, farm implements, fertilisers, and pesticides, to increase farm productivity. This SDG also promotes international cooperation to increase farmers' access to techniques for increasing productivity.

Goal 3: Good Health And Well-Being

The world is a better place to live now because of improved medical facilities. Life expectancy has increased, and infant and maternal mortality have decreased drastically. Most of the life-threatening diseases, like smallpox, are eradicated, and others, like HIV, are well under control.

For any country to develop and prosper, we need healthy citizens. The United Nations has underlined its importance by making it its goal. This SDG aims to provide good nutrition and medical facilities for people to overcome various communicable and non-communicable diseases. Health and good nutrition are the aims of this SDG which cannot be achieved without alleviating poverty hence it also shows the interconnectedness of SDG goals here.

Countries are making progress and rapid improvements in health conditions, but simultaneously, some countries need to catch up as they have a low life expectancy and high infant mortality. Work also needs to be done to equalise the efforts of the government to achieve SDG so that all sections of society, whether rich or poor, are gaining the same benefits from the schemes and plans initiated by the government to achieve good health and well-being. The gender equality goal also comes into play when different sections need healthcare.

Goal 4: Quality Education

The literacy rate has gone up since 2000, leading to rapid strides towards the target of universal primary education. The net attendance ratio has increased overall, promising a more educated society. The headcount of girls is also increasing in the schools. These are very positive indicators towards the goal of education for all.

It is understood that without education, we cannot expect our society to become efficient, and thus education promises a developed and stable future. Therefore, it has been made one of the SDGs. At times, working on this SDG becomes difficult due to wars, emergencies, epidemics, or famines. Recent wars like the Russo-Ukrainian War and the Israel-Gaza War have led to a decrease in the number of students, but efforts are still ongoing to provide students with the highest quality education.

This ensures that all girls and boys complete free primary and secondary schooling by 2030.

Children from poor households are generally devoid of educational opportunities. The SDG aims to provide quality vocational training to pupils, irrespective of gender and economic status, leading to universal access to higher and quality education. This SDG automatically interrelates with the Gender Equality SDG, as everyone will have to integrate towards the completion of this goal.

Goal 5: Gender Equality

Time and again, countries have shown that the more empowered their women are, the more that country flourishes. Empowering women improves the overall quality of life for a family. An educated woman, in turn, educates the entire family and becomes the reason for the prosperity of the upcoming generation. Empowered women direct their families to a safer position, and the family unit is more stable than the one where women remain dependent.

UNDP has encircled gender equality as its goal, and there has been remarkable progress in the past 20 years. With the initiatives of authorities and the support of societal factors, more girls are enrolling in school and higher education. There are more girls at different workplaces than ever. Despite the change, the discrimination against women in certain areas is very high, and they are still paid less in certain professions. They are systematically denied equal work or equal advantages. During times of natural disaster, migration, and war, women and children are still the most affected. This calls for much more work that needs to be done in this direction to improve the quality of life for each gender.

Thus, it is crucial to give women rights to land, property, and sexual and reproductive rights. Modern men should be imparted the same knowledge so that they understand the rights of women as indispensable and simultaneously fulfill their duty to bring about a harmonious change in their mindsets.

Goal 6: Clean Water And Sanitation

Potable water is a very scarce resource. It is alarming how most regions of the world are facing water scarcity and regular droughts. Dwindling freshwater resources are a matter of concern for every region of the world. Desertification is increasing day by day, and it is projected that 25 percent of people in the world will face water shortages on a regular basis by 2050. Discarding waste in water bodies has made the water sources vulnerable, and it is vital to save them by any means. We must invest in achieving universally safe and affordable drinking water so that the world population has a means of survival.

A 2015 UN report says that 4.5 billion people lacked safely managed sanitation services (with adequately disposed or treated excreta), and 2.3 billion lacked even basic sanitation. Safe and affordable drinking water for all by 2030 requires investing in adequate infrastructure, providing sanitation facilities, and encouraging hygiene. Thus, ensuring safe and hygienic sanitation will reduce the wastage of other resources, which are important for survival.

Goal 7: Affordable And Clean Energy

There is a decrease in the percentage of people without electricity. The number of people not enjoying energy supplies has gone down too. The prime concern is that with the increasing population, the demand for cheap energy supplies is increasing, and over-dependence on fossil fuels for energy generation is posing a grave danger to the climate. To counter these issues, we have to opt for clean energy resources that are affordable both for our pockets and for the environment.

Energy resources like solar, wind, tide, and geothermal are clean sources of energy that generate electricity without much harm to the surroundings. Therefore, SDG 7 encourages us to work on developing resources that are vital for environmental parameters. Under this, we aim to achieve a cleaner and more sustainable atmosphere and fulfil energy requirements by 2030.

Goal 8: Decent Work And Economic Growth

The trend in world poverty has shown a decrease in the number of people living below the poverty line over the past 25 years. Despite the global economic crises of 2008 and the COVID-19 pandemic of 2020–21, there has been an increase in the middle class from 1991 to 2015.

However, widening inequalities and a lack of jobs have caused distress in the labour force. According to the International Labour Organisation, more than 204 million people were unemployed in 2015. The SDGs promote innovation that leads to new economic opportunities and sustained growth in the job market and for the nation.

SDG 8 aims for full employment for all people, including young people, untrained labourers, and people with disabilities, by 2030. Everyone should receive equal pay for equal work. It is aimed at fully and efficiently utilising all productive resources for the nation's growth. The SDGs also focus on eliminating poverty, bonded labour, and human trafficking, as well as increasing job opportunities, productivity, and decent work.

Goal 9: Industry, Innovation And Infrastructure

EEconomic growth is critically dependent on investment in infrastructure, research, and development. Technological progress, with an ecosystem in mind, can only be achieved through safe and scientific innovation. Renewable energy resources are the only feasible solution, given the increasing population, to control environmental degradation. Transportation and communication technologies also require advancements to meet the needs of the upcoming generation.

Innovation promotes efficient energy usage. Scientific research will drive sustainable development by introducing new industries that align with environmental needs. Sustainable industrialisation is crucial, as it increases the industrial share of employment for the youth, with a progressive approach for the nation. SDG is focused on implementing cleaner infrastructure and technologies, with all countries collaborating based on their capabilities.

Goal 10: Reduced Inequalities

Income inequalities are on the rise despite efforts from countries and their respective governments. The gap between the richest and poorest countries is increasing systematically. According to the World Economic Forum, the richest countries (10%) hold 76% of the entire world's wealth, whereas the poorest countries only have a 2% share of global wealth. Factors like technological progress, globalisation, and commodity price cycles have exacerbated economic disparities within countries.

Regions in Africa and America face the highest income disparity, while Europe experiences the least disparity among countries. These widening disparities require sound policies to empower lower-income earners and promote economic inclusion for all, regardless of sex, race, or ethnicity.

SDG 10 has been designed to address this issue, reduce the gap between low- and high-income groups, and promote inclusiveness across all social and political spheres of society. This goal advocates for inclusiveness in the social and political arenas of the country for all sections of society.

Internationally, fiscal policies should aim to ensure equal distribution of benefits from economic outputs among those who generate them.

Goal 11: Sustainable cities and communities

Our society is moving towards urbanisation, so we need to plan sustainable cities for the future that are full of urban spaces. Cities are becoming megacities, with population booms in developing countries. We see slums and poor neighbourhoods within the vicinity of these high and well-developed urban spaces.

In order to make cities sustainable, we have to work upon building cities that are greener, have open public spaces for leisure activities and cleaner transport. The cities should also have more career options, affordable housing and resilient structures that brings in prosperity in an inclusive manner for all income and societal groups. It should have structures that may avoid critical loss to humans lives and property and are able to mitigate any emergency situation.

Goal 12: Responsible Consumption And Production

The world requires urgent action for the responsible consumption and production because human population is increasing everday and in order to maintain this population, we need to have food. The food requires input in the form of fertilisers and water. Fertilisers though increase the production but cause harm to the environment. Water, an already scarce resource is overutilised and needs to be saved from further exploitation in quality and quantity. The wastage of food is actually not a feasible action any more. So we need to be more responsible while consumption of food and resources used for its production. Therefore, overall, this SDG aims to improve both the production and consumption patterns of all commodities by 2030 to establish a more sustainable future for upcoming generations.

Other resources are also being exploited at an unexpectedly fast rate due to the global demand. Inefficient production processes for developing human wants and desires have led to wastage to existing supplies of raw materials. Thus overall, this SDG aims to improve both the production and consumption patterns of all commodities by 2030 in order to establish a more sustainable future for upcoming generations.

Goal 13: Climate Action

Pollution and the unrestrained burning of fossil fuels are the major factors causing climate change worldwide. Increasing greenhouse gas emissions are raising the earth's temperature every day. Greenhouse gas emissions are more than 50 percent higher than in 1990. Toxic products from industrial and vehicle pollution have also contributed to environmental changes, which are increasingly unsuitable for the livelihoods of future generations. The irreversible damage caused by plastics and nuclear wastes needs immediate attention.

Rising temperatures and climate change result in numerous natural disasters that lead to billions of dollars in losses worldwide. Therefore, this goal aims to mobilise international and national resources to mitigate environmental damage and work towards its possible recovery. The SDG will require political, economic, and human resources to bring about the much-needed change for the world.

GOAL 14: Life Below Water

Oceans are very important to the mankind as they make the planet habitable for humans and other forms of life. It is a vital resource which provides numerous other resources. We have to manage this resource in order to maintain the supplies of all other resources .

Coastal biodiversity is used by three billion people for their livelihood and the way we are consuming ocean resources, it is reaching unsustainable level. So, SDG 14 aims to bring sustainable yields back to the oceans and prevent its over-exploitation as well as address the impacts of ocean acidification. Oceans are absorbers of carbon dioxide as well so we must work upon their saving as they control a green house gas. So protecting oceans is directly or indirectly protecting us.

GOAL 15: Life On Land

Humans depend on earthly gifts as much as they depend on the resources of the oceans. Life sustains with the help of forests, atmosphere, land, and the creatures that inhabit them. Forests are natural suppliers of fresh water and maintain ecological balance, crucial for sustaining life. The decrease in forests contributes to the expansion of desert areas, known as desertification.

Statistically, 13 million hectares of forests are lost or destroyed every year. Even with legislation aimed at conserving natural forests, species continue to decline. Wildlife trafficking further threatens animal life, posing risks to human survival.

Therefore, SDG 15 encourages universal action to address these challenges exacerbated by changing lifestyles. This will secure life-supporting factors such as food, water, atmosphere, land, and climate, ultimately promoting a healthy and peaceful life for all.

Goal 16: Peace, Justice And Strong Institutions

We cannot hope for sustainable development without peace, stability, human rights, and effective governance based on the rule of law. Yet, our world is increasingly divided. We want peace and justice for humanity and are looking for strong institutions to achieve this. Some regions fall into an endless cycle of conflict and violence, while others are peaceful and prosperous. Therefore, all countries must take appropriate actions to combat these situations and foster the much-needed growth of a sustainable society that encourages positive change for future generations.

SDG 16 aims to reduce all forms of injustice and violence and promote a world governed by laws and order, resulting in institutions that govern people uniformly and in the most productive manner.

Goal 17: Partnerships For The Goals

Realisation of all SDGs is only possible when all countries collaborate and put their efforts to the nth degree, to bring about the changes we desire for a sustainable, peaceful, progressive, and liveable world. Since the world is more interconnected than ever, it is now mandatory to join hands with other countries and work towards our common goals.

Improving access to technology and knowledge with ever-improving scientific means will take our initiatives and ideas further. Coordinating such activities with developing nations will bring about the changes needed by the world. We need improvements in debt situations, pollution levels, equality, and partnerships to make this planet sustainable for future generations. So, let's join hands from east to west and north to south to make the world a fair and peaceful place to live.

Ai Reboot

1. Which SDG is working on to reduce violence, injustice and promote a world that works under laws and orders?

2. Which SDG is aimed at combating the processes that lead to the wastage of resources and the depletion of existing supplies of raw materials on Earth?

Brainy Fact

The concept of SDGs originated at the Rio+20 Summit in 2012, the largest summit in UN history. Colombia and Guatemala suggested establishing new goals to succeed the Millennium Development Goals, which were initiated in 2000 with the aim of halving poverty by 2015. This objective has been achieved, with poverty, defined as living on less than $1.25 a day, reduced by half. Setting such goals proves effective, as they help organisations and countries align their agendas and prioritise funding in a complex world.

The new goals resulted from a three-year process that included 83 national surveys and engaged over 7 million people, marking the most extensive consultation in UN history. A report indicates that Sweden is most likely to achieve these goals first, with Norway, Denmark, Finland, and Switzerland following closely behind.

Understanding the Capstone Project

Having studied the various facets of AI, you can create your own small version of an AI model that can help to solve a real-life problem. This type of AI model is called a capstone project. A capstone project is a comprehensive, independent, and final project undertaken as a part of the curriculum designed to assess the skills, knowledge, and expertise a student has acquired. Such a project often involves researching a topic, evaluating a new technique or method, developing a health plan, researching a character or event in history, or even the composition of a sketch or play.

No matter what type of project you choose to undertake, the result is the same. You can demonstrate your understanding of the course material and your readiness to enter the professional world to kickstart your career. It's a rewarding experience if it's done well.

So, a capstone project provides an opportunity for students to integrate all the knowledge learnt during the course and demonstrate it. Ideally, a great capstone project is one that:

- You have great interest and passion for
- Has practical usage/value
- Doable in a particular timeframe
- Contemporary
- And helps in your career advancement.

Digital Literacy

Video Session

Scan the QR code or visit the following link to watch the video: What is a Capstone Project?
https://www.youtube.com/watch?v=yBs2Vb5I1f54
After watching the video, answer the following question:
Define Capstone Project? What is its purpose?

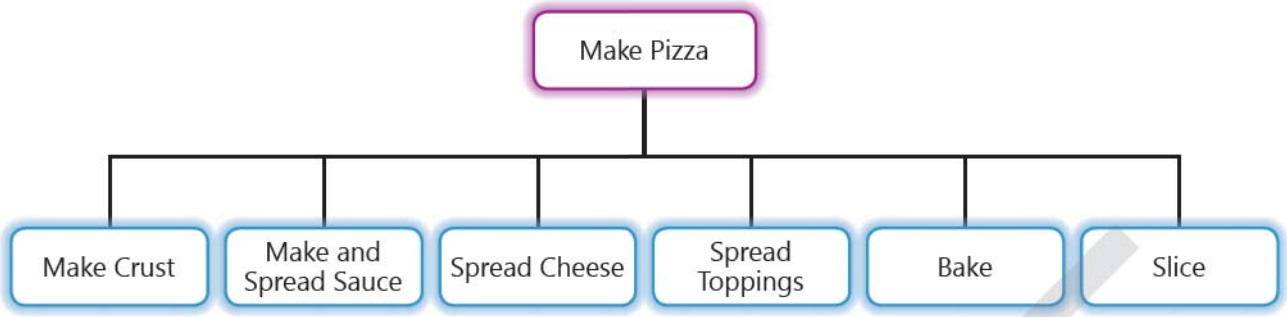
Some of the examples of capstone projects in AI that you can develop are as follows:

- Studying images to diagnose diseases,
- Forecasting student results,
- Creating a chatbot for the school admin department or counsellor to handle parents'/students' queries using IBM Watson or Google Dialogflow,
- Image Classifier,
- Analysing social media to assess emotions,
- Using regression to predict a trend.

Problem Decomposition

Solving real-life problems is complicated. During coding, we follow problem decomposition methodology that can be applied to real-life problems as well. Here, we break down a complex problem into a series of small, more manageable problems.

The process of making pizza can be split into sub-problems:



Steps for problem decomposition are as follows:

1. Understand the problem and express the problem in your own words:
   - Understand the required inputs and outputs.
   - Ask questions for clarity (in class, these questions may be directed to the teacher, however, you can also ask yourself or your colleagues).
2. Break down the problem into several big parts. Write them down on a paper.
3. Divide any larger complicated part into smaller parts. Continue this until all parts are small.
4. Code the smaller parts one by one. Use the following methodology:
   - Analyse how to implement the code.
   - Write the code/query.
   - Test each code individually.
   - Fix the problem(s), if any.

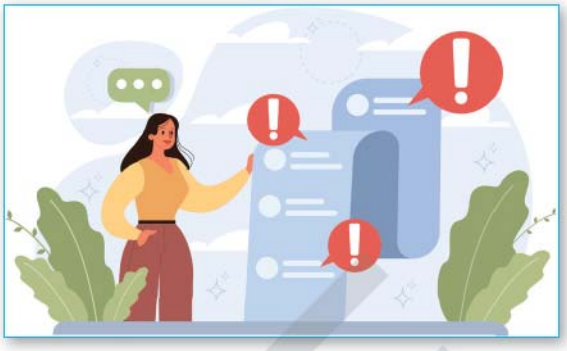
Imagine you want to create your first website. How would you decompose this task? Think about the following while decomposing:

- What would be the name of the website?
- What colour combination can be used?
- How many web pages are to be included on the website?
- Who will be the target audience?
- What kind of images are to be posted on the website?
- What video/audio would be included?
- Which software is to be used for website development?
- How will you test your website?
- Debug and make improvements.

Defining the Problem

Nowadays, Artificial Intelligence is one of the most transformative technologies. Every AI project goes through these six steps:

Understanding the problem



The beginning of any project requires describing the problem, which is done in the problem-defining phase of the project life cycle. It is a crucial stage that requires in-depth research of the problem so that the desired solution is well-written in an absolute understandable form. This helps progressively in all the following stages of the project life cycle, and is also used to trace back any missed out features that was initially planned to implement in the AI model. Successful implementation of the AI model is critically dependent on this stage of the AI life cycle.

Data collection

Once the problem is identified and defined, we then begin to collect the data for it. The data that needs to be collected is identified and recorded using machines that record real-world signals for AI model to work successfully. Data acquisition is followed by data cleansing, that drops the irrelevant information with respect to the problem at hand. The inputs are then digitised for processing by the AI model in place.

Defining the features

In order to define the data values gathered, we must be clear with the heads of the data, its data types, and about the missed and duplicate values that might be encountered while recording data. If gaps and outliers are recorded, then understand how much variation in the AI model is made to withstand without altering the final result of the model. So, defining the data generally involves making it more suitable for analysis and processing for the AI project that follows.

AI Modelling

This is the phase of implementation of the AI model using suitable platforms, programming languages, and constructs. An AI model is a program that has been trained on a set of data to recognise certain patterns or make certain decisions without further human intervention. They apply different algorithms to relevant data inputs to achieve the tasks, or output, they've been programmed for. This model is developed to solve the problem defined in the first phase. The model is then trained with the training data. The training is iterative in nature, so that the system is prepared for the most unexpected scenarios. This produces a refined model ready to be evaluated in the upcoming phase.

Evaluation

The designed and tested model is now to be checked for its real-world execution. The AI model is now given test cases and test data to check to see if it works accurately in real life and uncovers the errors that might have been left hidden in the modelling and training phase. The trained and tested model is then reformed based on the unprecedented outputs that are encountered during evaluation phase. Sometimes the defined parameters need tunings and adjustments to suit the problem in hand.

Deployment

After the modelling and evaluation phase, the AI model is deployed or implemented in real-life scenarios. The model is then integrated with existing systems. The new application is then utilised and upgraded as and when some unprecedented scenarios cause some undesirable results. The model could be working dynamically in an online environment as well as offline, like reporting things to the manager.

Understanding the nature of a problem provides insight into the components and attributes of the yet-to-be-implemented solutions. A good understanding of a problem guides future decisions to make at the later project stages, especially decisions such as determining if an AI solution is even feasible. At the core of every AI model is "finding patterns in data". If the data shows no patterns, then most probably, the problem cannot be solved using AI.

A successful problem-defining process requires a basic analysis and evaluation of the project-related problems, their reasons and methods. Finding the right problem definition is usually an iterative process. It can reveal more questions and points to consider that would have been ignored without the problem definition process. The questions below serve as a reference point for a thorough analysis of the problem and the problems that surround it. Just spending time answering the following questions can save you weeks and months working on problems that proved impossible for previously unknown reasons:

- What is the problem that needs to be resolved?
- Why do you need a solution to your problem?
- How should we work on the solution to the problem?
- Which aspects of the problem does the AI model solve?
- How do I need to interact with the solution to the problem?
- Which category of data will be involved? (Classification)
- How much or how many? (Regression)
- Can the data be grouped? (Clustering)
- Is there any unusual pattern in the data? (Anomaly Detection)
- Which option should be given to the customer? (Recommendation)

It is essential to determine which of these questions you're asking and in what way answering them helps solve your problem.

Ai Task

Divide the class into the group of 10 students.

Think of an automated security system for a school. Define the problem, i.e. the features you think are required. Mentioned the data set required. Design a dry run of the model using a one-act play and find out if the system, if executed, is suitable for the problem. Discuss your findings with other groups.

Sample Capstone Projects

- Rule-based chatbot on IBM Watson or Google Dialog flow or joonbot.com/templates/.
- Use the dataset given on https://www.kaggle.com/camnugent/california-housing-prices. Use only median_income (x-value) and median_house_value (y-value) columns for plotting a scatterplot. Use this scatterplot to calculate the Regression Line of best fit. Then use the line to calculate y for a new value of x. You can do this using Excel or Python.

Similar datasets can be found at:

https://www.telusinternational.com/articles/10-open-datasets-for-linear-regression

Additional Study Material (Using Python to create Regression Model)

- Project 7. Car Price Prediction using Machine Learning with Python | Machine Learning Projects https://www.youtube.com/watch?v=L3OtLaCbJC8
- Project 8. Machine Learning in Python: Building a Linear Regression Model https://www.youtube.com/watch?v=R15LjD8aCzc

Sample Videos of Capstone Projects prepared by students

- Project 1. AI Capstone Project Video on Heart Disease Predictive System https://bit.ly/3Xyjehx
- Project 2. Social Distancing Monitoring https://bit.ly/3CPgT9T
- Project 3. Project Happify http://bit.ly/3IIV53z

Access sample logbooks of capstone projects developed by students through this link https://drive.google.com/drive/folders/1xqu7pWvqQehw-ZI3fupVvQywNslQTzMx?usp=sharing

Ai Reboot

Suppose you are creating a mobile app for grade 11 students. The app will solve user queries on 'Mental Health'. Decompose the creation of the mobile app into several simple steps.

At a Glance

- Design thinking is a process to solve problems creatively.
- Design thinking consists of five different stages: Empathise, define, ideate, prototype and test.
- Empathy Map enables designers to understand, empathise, and, share user feelings.
- Empathy Maps provide an overview of the user as a whole and are not sequential depicted by four quadrants how a user says, thinks, does and feels.
- The SDG goals are interrelated as they affect each other and one cannot be completely achieved without incorporating the other.
- Capstone project provides an opportunity for the students to integrate all the knowledge learnt during the course and demonstrate it.
- AI project goes through the six steps: defining the problem, data collection, defining the features, AI modelling, evaluation, and deployment.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. Which of the following is not one of the five different stages of design thinking?
   a. Empathise ◯ b. Ideate ◯
   c. Evaluate ◯ d. Protoype ◯

2. How do "Individual Brainstorming" help introverts?
   a. They can express their ideas in writing ◯ b. They can voice their thoughts ◯
   c. They learn to stand for themselves ◯ d. They get to know others ◯

3. What is the most important part of Empathy Map?
   a. Says ◯ b. Think ◯
   c. Does ◯ d. User ◯

4. In what situation the problem cannot be solved by AI?
   a. If we have a lot of data ◯ b. If we have little data ◯
   c. If we have many patterns in data ◯ d. If we have no pattern in data ◯

5. Which phase involves identifying and recording real-world signals for AI models?
   a. Defining the features ◯ b. Data collection ◯
   c. AI Modelling ◯ d. Deployment ◯

6. What is crucial for a successful problem-defining process?
   a. Iterative analysis and evaluation ◯ b. Data cleansing ◯
   c. Defining the features ◯ d. AI Modelling ◯

7. What does Evaluation not do?

a. Understanding the nature of the problem. ◯ b. It uncovers errors that were left. ◯

c. Defined parameters are tuned. ◯ d. Model is checked for real world execution. ◯

8. To gather relevant information from customers, a simple and efficient method called _____, may be used.

a. 4W ◯ b. 5W ◯

c. 4W and 1H ◯ d. 5W and 1 H ◯

B. Fill in the blanks.

1. Design thinking is _____, which means that it uses evidence of how consumers (people) interact with a product or service, rather than how organisations think they will handle it.
2. A _____ is a simple experimental model for a proposed solution.
3. The traditional Empathy Map is divided into _____.
4. The Sustainable Development Goals (SDGs) are also termed as the _____ .
5. _____ and _____ have been the most successful of all, in uplifting the people out of poverty.
6. The United Nations has considered food scarcity and undernourishment of children and has made _____ as one of its goals.
7. _____ aims for full employment for all the people including young people, untrained labour and people with disabilities.
8. _____ are the only feasible way to go, with the increasing population, to keep the environmental degradation under control.
9. Water an already scarce resource is overutilised and needs to be saved from further exploitation in _____ and _____ .
10. _____ products of industrial and vehicle pollution have also led to the change in the environment.

C. State whether the following statement is True or False.

1. Ideate is the first step of design thinking. _____
2. In Empathy Map, we try to move out from user's perspective and view the problem from an observer point of view. _____
3. The Sustainable Development Goals were adopted in 2017 in New York. _____
4. Wars make it difficult to implement education for all. _____
5. Women are always given equal opportunities to educate and work for their advantage. _____
6. In order to remove inequalities we should be ready to mitigate any emergency situation. _____
7. Pollution and unrestrained burning of fossil fuels are the major factors for causing the change in climate worldwide. _____
8. Statistically 13 million hectares of forests are lost or destroyed every year. _____
9. Data acquisition is followed by data cleansing that drops the irrelevant information. _____
10. AI model training is a one step process. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. What is Design Thinking?

Ans. Design thinking is a process to solve problems creatively. It solves problems by putting consumers' needs first. So, design thinking is "people-oriented", which means that it uses evidence of how consumers (people) interact with a product or service, rather than how organisations think they will handle it.

2. What do W and H represent in 5W and 1H?

Ans.
- WHO (stakeholders)
- WHAT (issue, problem, need)
- WHEN (time)
- WHERE (location/situation)
- WHY (reason)
- HOW (solution will solve/improve problem)

3. What is an Empathy Map? How does it help?

Ans. An Empathy Map is a collaborative visualisation used to clarify our understanding of a specific type of user.

It helps:
- To create a common understanding of user needs
- In decision-making. It also helps in deepen understanding and gaining insight into user behaviour.

4. What is SDG Goal 10?

Ans. Income inequalities are on the rise despite all efforts from the countries and their respective governments. The gap between the richest and poorest countries is increasing systematically. According to world economic forum, the richest countries(10%) hold 76% of the entire world wealth whereas the poorest countries only have a share of 2% of the entire wealth.

5. What is problem decomposition?

Ans. Complex problems cannot be solved in one go. So, we try to simplify the complex problem to analyze it. This is known as Problem Decomposition. Here, we break down a complex problem into a series of small, more manageable problems.

B. Long answer type questions.

1. What does four quadrants in the Empathy Map imply?

Ans. The traditional Empathy Map is divided into 4 quadrants (Says, Thinks, Does, and Feels), with the user or person in the middle. Empathy Maps provide an overview of the user as a whole and are not sequential.
- The Says quadrant contains what the user says aloud in an interview.
- The Thinks quadrant contains what the user is thinking about his/her experience.
- The Does quadrant depicts the actions of the user.
- The Feels quadrant denotes the user's emotional state, often represented as an adjective along with a short sentence for context.

2. Elaborate the five different stages of design thinking.

The five different stages of design thinking are given below:

Ans.
- **Empathise:** In the first stage, designers observe consumers to gain a deeper understanding of how they interact with or are affected by the product or problem. Observations must be made with empathy, that is, without making any judgements or conveying any preconceived notions about consumer needs. Empathy observation is powerful because it can uncover issues the consumers didn't even know they had or that they could not verbalise. From this point on, it is easier to understand human needs.
- **Define:** In the second stage, you collect the observations from the first stage to define the problem to be solved. Think about the difficulties your consumers may face, the problems they face again and again, and what you have learned from how they were affected by the problems. Once you have summarised your results, you can define the problem statement.
- **Ideate:** The next step is to gather ideas to solve the problem you identified. These brainstorming sessions can be conducted in a small group, and your team can meet in an office area that encourages creativity and collaboration, an innovation laboratory, or the team members can work alone. The important thing is to generate many different ideas. At the end of this process, you will have some ideas to move forward.

- **Prototype:** This is the stage where an idea becomes a real solution. A prototype is a simple experimental model for a proposed solution. The prototype may not be perfect. The focus of the prototype is to quickly come up with a concrete version of the idea and see how it will be accepted by consumers. Examples of prototypes include a website to test consumer demand for products, a 3D printed product, etc.
- **Test:** Once you have made prototype solution available to consumers, you need to observe how they interact with it. During this testing phase, you collect feedback about your work.

  The design thinking process is iterative. At the end of the fifth stage, you may have to return to one or more other stages. Perhaps the test indicates that you need to develop another prototype, so, you will return to the stage. Or maybe it turns out that you misunderstood the needs of consumers, then, you need to go back to the early stages of the process.

3. What is the purpose of the Ideation stage?

Ans.
- Ask the right questions and innovate, focus on your users, their needs, and your understanding of them.
- Combine the views and strengths of team members.
- Get rid of the obvious solutions in your mind and let the team surpass them.

4. What do you mean by prototype?

Ans. A prototype is a simple experimental model for a proposed solution. The prototype may not be perfect. The focus of the prototype is to quickly come up with a concrete version of the idea and see how it will be accepted by consumers. Examples of prototypes include a website to test consumer demand for products, a 3D printed product, etc.

5. Differentiate between brainstorming, brain dump, and brainwriting.

Ans.

Brainstorming	Brain dump	Brainwriting
Vocal process	Silent process	Some team members are vocal and some are silent.
Few members can dominate the session.	Encourages uniform participation.	Encourages uniform participation.

C. Competency-based/Application-based questions:

1. After learning the importance of design thinking process in the present era of model making, Neerja was tempted to avoid the prototyping phase in order to do the same during AI project development. Is she right? Can we skip prototyping stage of design thinking? Yes or No. Please Elaborate.

Ans. We cannot go ahead without prototyping because a prototype gives a hint of the actual product that the designer intends to build. It also gets us the first glimpse of the AI model that will be functional and how it will be used by consumers. The concrete version of the idea of the designer and AI developer is shown in the prototype which builds a solid pathway for the development to proceed ahead.

2. The term "sustainable development" was indeed coined several decades ago, gaining widespread recognition with the publication of the Brundtland Report, also known as "Our Common Future," by the World Commission on Environment and Development in 1987. This means the Term "Sustainable Development" have been around since few decades but the Sustainable Development Goals are planned now. Why?

Ans. While the concept of sustainable development has been around for decades, the formulation of specific goals like the SDGs required a combination of evolving understanding, global collaboration, scientific advancements, policy

development, increased awareness, and learning from past initiatives. After a suitable framework was developed the SDGs were put in practice with clear intentions and final results for developed as well as developing nations.

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. Which of the following best describes the "empathise" stage in the design thinking process?
   a. Testing prototype solutions ◯
   b. Defining the problem statement ◯
   c. Observing consumers without judgement to understand their needs ◯
   d. Brainstorming possible solutions ◯

2. What is the main goal of the "Ideate" stage?
   a. To create a working model of the solution ◯ b. To understand consumer needs ◯
   c. To generate a wide range of ideas ◯ d. To test the prototype with consumers ◯

3. Why is the testing stage important in design thinking?
   a. To finalise the design ◯
   b. To collect feedback on the prototype from consumers ◯
   c. To launch the product in the market ◯
   d. To brainstorm new ideas ◯

4. In design thinking, why is empathy important in the first stage?
   a. It helps in creating technical specifications ◯
   b. It allows designers to impose their own ideas ◯
   c. It uncovers issues consumers may not realise they have ◯
   d. It defines the solution immediately ◯

5. What does the "Says" quadrant in an Empathy Map represent?
   a. The user's emotional state ◯ b. The actions of the user ◯
   c. What the user says aloud in an interview ◯ d. What the user is thinking ◯

6. Which quadrant of the Empathy Map would capture the user's actions?
   a. Says ◯ b. Thinks ◯
   c. Does ◯ d. Feels ◯

7. Which of the following regions has seen a decrease in extreme poverty?
   a. Sub-Saharan Africa ◯ b. East Asian Countries ◯
   c. Central America ◯ d. South America ◯

8. What is a key factor for a country's development and prosperity according to Goal 3?
   a. Economic growth ◯ b. Healthy citizens ◯
   c. Political stability ◯ d. Technological advancement ◯

B. Fill in the blanks.

1. Children from poor household are generally devoid of educational _____ .
2. An _____ in turn educates the entire family and become the reason for prosperity of upcoming generation.
3. It is projected that _____ percent of people of the world will face water shortage on regular basis by 2050.
4. The prime concern is that with the increasing _____ the demand for cheap energy supplies is increasing and _____ on fossil fuels for the energy generation is posing a grave danger for the climate.
5. _____ is used by three billion people for their livelihood and the way we are consuming ocean resources it is reaching an unsustainable level.
6. _____ is a crucial stage that requires in-depth research of the problem so that the desired solution is well written in an absolute understandable form.
7. In Evaluation, sometimes the defined parameters need _____ and _____ to suit the problem in hand.
8. _____ and statistical models are used to forecast the potential outcome.

C. State whether the following statement is True or False.

1. Ideate stage of the Design Thinking involves brainstorming. _____
2. Praying is a quadrant of Empathy Map. _____
3. The SDG goals are exclusive of each other. _____
4. SDG 3 works on eradicating life threatening diseases. _____
5. The SDG aims to provide quality vocational training to the pupils keeping in mind gender and economic status. _____
6. Energy resources like Solar, Wind, Tidal, Geothermal are polluting sources of energy. _____
7. SDG 9 is directed to bring in cleaner infrastructure and technologies with all countries together taking actions based on their capability. _____
8. Regions of Africa and America face the highest income disparity, whilst Europe is facing the least disparity within the countries. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. What is the purpose of Design Thinking?
2. What is the difference between Evaluation and Deployment in an AI project?
3. Why are SDG 14 and 15 important?
4. What is AI modeling?
5. Define capstone project.

B. Long answer type questions.

1. What do you understand by "Deployment"?
2. What is meant by evaluation of an AI model? Why is it important?

3. Why is brainstorming very popular with design teams?
4. What all is done in the "Empathy" stage?
5. "Identifying a problem is more important than providing a solution." Do you agree? Why/Why not? Give examples.

C. Competency-based/Application-based questions:

1. In a company ABC, Design Thinking was introduced and employees were asked to employ it before actual beginning of any project. Anuj was confused as to which stage of the design thinking requires the most resources and what should he work the hardest on, so that a practically implementable prototype is developed and the time consumption during actual project making would be minimised? What would be your response to this query and discuss your answer for Anuj's easy understanding.
2. In a city, a team of data scientists started on a mission to build an AI system to predict traffic patterns. They knew their success was based on collecting the right data and defining key features. Without accurate traffic data and well-defined variables like time of day, weather, and road conditions, their AI would struggle to make accurate predictions. This careful groundwork was crucial to avoid creating an unreliable system that could mislead drivers and city planners, highlighting the essential role of data collection and feature definition in AI projects. List the factors which need to be ascertained for the same.

AI In Life

Environmental Awareness

Some SDGs are important for both developed and developing nations. Find out which SDGs are these. Also, waste tyres can be linked to several SDGs. Find out which ones.

AI Deep Thinking

Creativity and Innovativeness

1. How can the integration of sustainable development principles into capstone projects across various academic disciplines contribute to the achievement of the United Nations Sustainable Development Goals (SDGs), particularly in addressing global challenges such as climate change, poverty, and inequality?
2. How can a capstone project that integrates multidisciplinary knowledge and demonstrates real-world applicability effectively bridge the gap between academic learning and professional practice, thereby enhancing a student's career prospects and contributing to societal needs?

Ai Lab

Experiential Learning

1. Use the online tool (https://online.visual-paradigm.com/diagrams/templates/empathy-map/empathy-map-template/) to create an Empathy Map for the following scenario:
   (Editor: Give QR code for the link above)
   - You are part of a hospital management team, tasked with improving the daily on going activities at the hospital. To ensure the smooth working, managing the daily log of patients, to meet the actual needs and concerns of patients, you will create an Empathy Map based on insights gathered from patient log, interviews, and observations.
   - You are a class XII student who wants to apply for an AI course overseas.
2. Identify a theme that can be classified as a social problem. Create a 5W and 1H problem canvas for it.
3. Using Design Thinking, give a solution to the problem "Improving the traffic congestion in early morning school hours".
4. Take an SDG Goal and brainstorm about your capstone project around that goal. You can start filling the logbook especially the following areas:
   a. Project Plan
   b. Minutes of the Meeting
   c. Problem Definition
   d. User Interviews

Answers

Exercise

A. 1. c 2. a 3. d 4. d 5. b 6. a 7. a 8. d

B. 1. people-oriented 2. prototype 3. 4 quadrants 4. Global Goals
5. China, India 6. ZERO HUNGER 7. SDG 8
8. Renewable Energy Resources 9. Quality and Quantity 10. Toxic

C. 1. False 2. False 3. False 4. True 5. False 6. False 7. True
8. True 9. True 10. False

Read the article Sustainable Water Solutions: The Role of AI in Reducing Water Wastage In India -

https://www.hindustantimes.com/cities/bengaluru-news/bengaluru-water-management-body-brings-in-ai-iot-technologies-to-manage-borewells-amid-crisis-report-101711431323128.html

Now answer the following questions:

1. Which is the estimated loss to the Indian GDP due to the ongoing water crisis?

2. What role can AI play in Addressing India's Water Crisis? Explain briefly in 2-3 lines.

3. How is AI helping in dealing with water management? Give 2-3 examples from the article.

4. Give examples from the article how several countries are using AI to tackle the water scarcity problem.

Also read India's own success story in handling water scarcity - Bengaluru water management body brings in AI, IoT technologies to manage borewells amid crisis -

https://www.hindustantimes.com/cities/bengaluru-news/bengaluru-water-management-body-brings-in-ai-iot-technologies-to-manage-borewells-amid-crisis-report-101711431323128.html

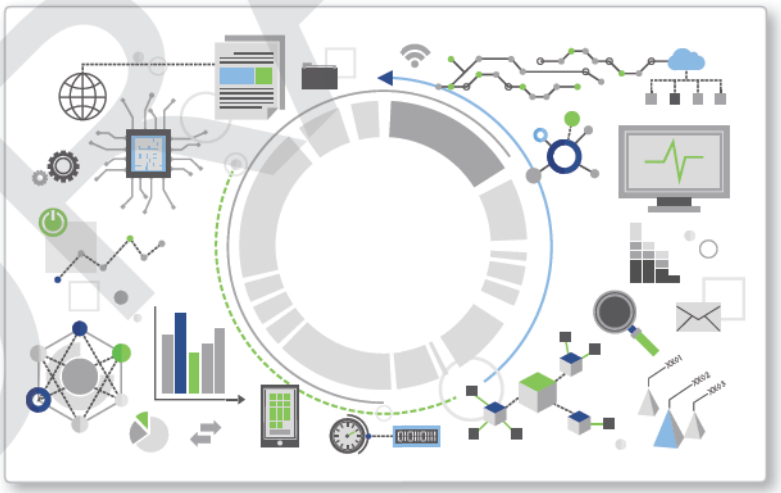
UNIT-5

DATA LITERACY–DATA COLLECTION TO DATA ANALYSIS

Learning Outcomes

- Data Literacy
- Exploring Data
- Calculating Measure of Central Tendency using Python
- Representation of Data
- Introduction to Matrices
- Operations on Matrices
- Data Preprocessing
- Data Collection
- Statistical Analysis of Data
- Variance and Standard Deviation
- Introduction to Matplotlib
- Order of Matrix
- Applications of Matrices in AI

Data can be described as a representation of facts or instructions about entities (such as students, schools, sports, businesses, animals, etc.) that can be processed or communicated by humans or machines. It is a well-known fact that Artificial Intelligence (AI) is fundamentally driven by data. AI involves transforming large volumes of raw data into actionable information that has practical value and can be utilised effectively.



Data Literacy

Data literacy involves the ability to find and use data proficiently. This encompasses skills like collecting data, organising it, checking its quality, analysing it, understanding the results, and using it ethically. Data can be **structured, semi-structured, or unstructured**. It should be properly collected, organised, and analysed to ensure that the input for AI models is valid and appropriate.

AI data analysis employs AI techniques and data science to enhance the processes of cleaning, inspecting, and modelling over both structured and unstructured data. The main goal is to extract valuable information that can aid in decision-making and drawing conclusions.

Data Collection

Data collection lets you record past events, so you can use data analysis to find patterns. From these patterns, you can build predictive models using machine learning to spot trends and predict future changes.

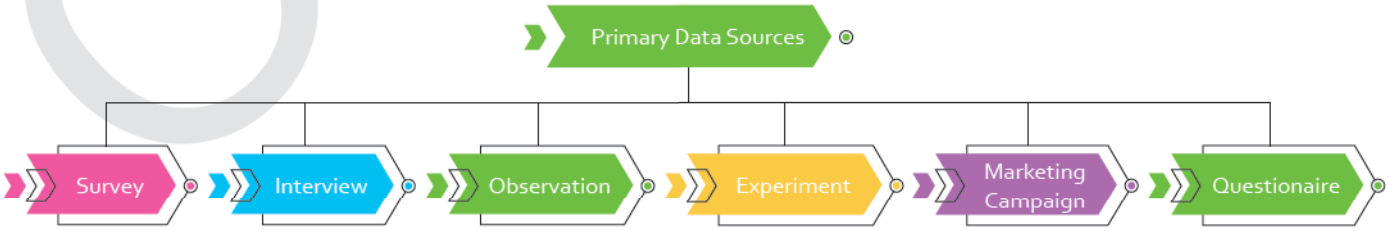
Data collection means gathering data from many sources, both offline and online. Collecting large amounts of data can be the hardest part of a machine learning project, especially on a large scale. The amount of data you need depends on the number of features in the data set. It's best to collect as much data as possible for accurate predictions. You can start with small batches of data and see how the model performs. It's important to collect diverse data to ensure your model covers various scenarios.

The **quantity of data** also depends on, how **complex your model** is. For simple tasks like, license plate detection, small batches of data might be enough. But for more complex tasks like, medical AI, you will need a lot of data. Before collecting data, data scientists need to understand the problem, the best solution, and the data requirements. Based on these requirements, they identify the data sources and collect the data. Data is essential for any project, and it is needed throughout the project's development. Therefore, identifying data needs, collecting data, and analysing it, is done repeatedly.

Data is the main source for data collection methods. In any type of research or company operations, collecting data serves primarily to support the identification of important variables, performance being the most important among them. Thus, the process of collecting data plays a vital role almost in every field. Depending on the kind of data being collected, there are two main categories of data collection methods: primary and secondary.

Primary Data Source

A primary data source refers to the original source from which data is collected firsthand. This data is obtained directly from its origin, without any intermediary sources or interpretations. Primary data sources include surveys, interviews, observations, experiments, and any other method where data is collected directly by the researcher or organisation for a specific purpose. This type of data is considered valuable because it is tailored as per the specific research or business needs and is often more accurate and relevant than secondary data, which is obtained from sources which have already interpreted or analysed the original data.



Let us study each source.

Source	Description	Example
Survey	Obtaining information from a population via phone, mail, questionnaires, or online forms. Useful for determining opinions, behaviours, and demographics.	An organisation conducts an online survey to gather employee's opinion on job satisfaction.
Interview	To obtain information, engage in direct connection with individuals or organisations. It might be structured, semi-structured, or unstructured.	Company conducts interviews with customers to gather feedback on products, understand customer needs, and identify market trends.
Observation	Observing and recording behaviour or events as they occur. Frequently employed in ethnographic studies, or when direct interaction is not feasible.	Observing consumer behaviour in stores or public places to gather insights on users' habits, preferences, and product usage.
Experiment	Manipulating factors to see how they affect outcomes. Used to determine cause-and-effect relationships.	Testing the efficiency of different advertising techniques on a group of people.
Marketing Campaign (using data)	Using customer data to forecast behaviour and improve advertising performance.	A company personalises email marketing initiatives based on previous client purchases.
Questionnaire	A specialised tool used in surveys – is a collection of questions meant to gather information from respondents. Can collect quantitative (numerical) or qualitative (descriptive) data.	For patient surveys to assess satisfaction, gather health histories, or monitor symptoms and treatment outcomes. A questionnaire may ask respondents to assess their happiness on a scale of 1 to 5 while also providing open-ended feedback.

Secondary Data Source

Secondary data sources provide information that has been gathered, processed, and published by someone else. This data can be obtained from a variety of sources, including research articles, books, reports, and internet databases, and is commonly utilised in research, analysis, and decision-making. Using secondary data sources might save time and resources because the data has already been collected and processed, but it is critical to assess the data's reliability and relevance to ensure it meets the research or analytical requirements.

Secondary data can be collected through the following methods:

Method	Description	Example
Social Media Data Tracking	Collecting information from social media sites such as user postings, comments, and interactions.	Analysing social media sentiment to understand audience reception of a new film release.
Web Scraping	Employing automated software to gather specific information and data from websites.	Scraping rental listings and prices from real estate websites for market analysis.

Method	Description	Example
Satellite Data Tracking	Collecting data on the Earth's surface and atmosphere through satellite.	Monitoring deforestation and land use changes using satellite imagery.
Online Data Platforms	Collecting data from websites that provide pre-compiled datasets for diverse purposes.	Kaggle, GitHub, KDNuggets, Google Dataset Search etc.
Books, textbooks, and encyclopedias	Information that has been researched, compiled, and written by authors or editors.	"Ancient Civilisations: A Comprehensive Guide" written by an expert historian.

Brainy Fact

Here are some fun facts about data:

- To download all of the information from the Internet would take 181 million years.
- Every two days, we generate as much data as we did from the beginning of time until 2003.
- There are almost as many digital information bits as there are stars in the universe.
- Less than 0.5 percent of the data we generate is ever used or evaluated.
- According to PragmaticWorks, poor-quality data costs global firms between 20 and 35 percent of their operating revenue.
- If you burn all the data created in one day onto DVDs, you could stack them on top of each other and reach the moon twice.

Exploring Data

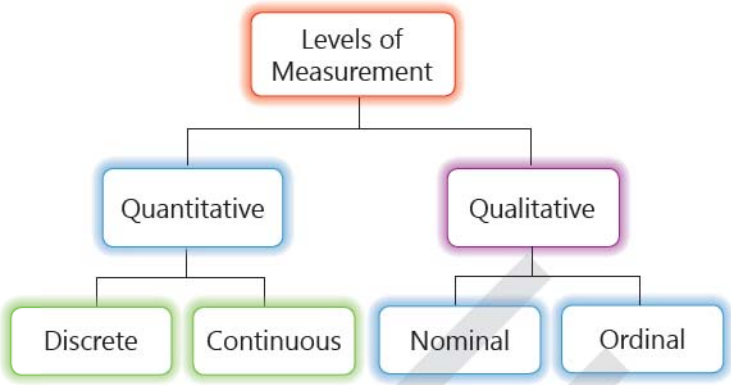
Exploration, one of the first steps in data preparation, is a way to get to know data before working with it. Exploring data involves familiarising oneself with the data, and understanding its value—whether it is usual, unusual, widely distributed, or extreme. This process not only helps in understanding the dataset better but also provides an opportunity to detect and rectify any data issues that might affect the analysis results. By addressing these problems during the exploration phase, one ensures that the conclusions drawn from the analysis are reliable and accurate. Data exploration use statistical methods and visualisation tools to:

- Evaluate the size and quality of your data.
- Detect outliers or anomalies.
- Identify possible links between data components, files, and tables.
- Look for similarity, patterns, relationships, and outliers.
- Determine the relationships between different variables.

Data exploration can be applied in a variety of areas, including banking, healthcare, retail, and marketing.

Levels of Measurement

The method used to measure a collection of data is known as the level of measurement. Not all data can be handled identically. It makes sense to classify data sets according to several criteria. Some are quantitative, others are qualitative. Some datasets are continuous, whereas others are discrete. Qualitative data might be either nominal or ordinal. Quantitative data can also be divided into two categories: interval and ratio.



The Four Levels of Measurement:

	Nominal	Ordinal	Interval	Ratio
Categorises and labels variables	✓	✓	✓	✓
Ranks categories in order		✓	✓	✓
Has known, equal intervals			✓	✓
Has a true or meaningful zero				✓

Nominal

In nominal measurement, the numerical values represent a unique "name" of the attribute. The cases may be ordered in any manner. For example, jersey numbers in cricket are measured at the nominal level. A player with the number 20 is not better than a player with the number 3 and is certainly not twice better whatever number 10 represents.

Nominal variables are like labels or categories—think car brands or seasons. They can't be ranked or used in calculations. Examples include eye colour, gender, or smartphone brands. Even if numbers are involved, like a player's jersey number, they're just identifiers, not for calculations or comparisons.

Examples:

What is your gender?
- (•) M – Male
- ( ) F – Female

What is your hair colour?
- (•) 1 – Brown
- ( ) 2 – Black
- ( ) 3 – Blonde
- ( ) 4 – Grey
- ( ) 5 – Other

Where do you live?
- (•) A – North of the equator
- ( ) B – South of the equator
- ( ) C – Neither in the international space station

Ordinal

In ordinal measurement, attributes can be ordered. The distances or intervals between attributes are irrelevant here. For example, in a survey, you can code educational qualification as, 0 = secondary; 1 = senior secondary; 2 = graduation; 3 = post-graduation; 4 = PhD. In this level of measurement, higher numbers mean more education. However, is the distance from 0 to 1 equal to 3 to 4? Of course, no. The interval between the values cannot be interpreted as an ordinal measure.

Ordinal data consists of categories arranged in a specific order, like rating a meal from "unpalatable" to "delicious." Although words, not numbers, are used, there's a clear progression from negative to positive. However, the actual difference between each category can't be measured. Like nominal data, ordinal data can't be used in calculations.

For instance, in a hotel survey, responses might be "excellent," "good," "satisfactory," and "unsatisfactory," ordered from best to worst. But again, the gaps between these ratings can't be quantified. Similarly, grading systems use letters to rank performance, but without additional context, it's impossible to determine the precise difference between grades.

Examples:

Grades:
A+ – Outstanding
A – Excellent
B+ – Very Good
B – Good
C – Fair
D – Needs Improvement

How do you feel today?
(•) 1 – Very Unhappy
( ) 2 – Unhappy
( ) 3 – OK
( ) 4 – Happy
( ) 5 – Very Happy

How satisfied are you with our service?
(•) 1 – Very Unsatisfied
( ) 2 – Somewhat Unsatisfied
( ) 3 – Neutral
( ) 4 – Somewhat Satisfied
( ) 5 – Very Satisfied

Interval

While measuring intervals, the distance between attributes is important. For example, if we measure temperature (in Fahrenheit), the distance between 30–40 is equal to the distance between 70–80. The interval between the values is interpretable. Interval level data can be used in calculations, but any comparisons cannot be done. 80°C is not four times hotter than 20°C (and 80° F is not four times hotter than 20°F). The ratio of 80:20 (or four to one) doesn't matter.

Interval level data shares similarities with ordinal data as it maintains a clear order, but it differs in that the differences between values can be measured. However, unlike ordinal data, interval scale data lacks a true zero point.

Examples:

Temperature

IQ score

Income ranges

Ratio

When measuring ratio, there is always an absolute zero point that makes sense. This means that you can use a ratio variable to construct a significant fraction (or ratio). Weight is a variable of proportion. In applied social research, most "number" variables are ratios, for example, the number of clients for a product. You can have zero customers and it makes sense to say, "we had twice as many customers last year compared to what we have this year."

Ratio scale data, is like interval scale data, features a true zero point and allows meaningful ratio calculations. For instance, consider the scores of a statistics final exam: 80, 68, 20, and 92 (out of 100). These scores can be ordered from lowest to highest (20, 68, 80, 92) or vice versa, with the differences between scores holding significance. For example, 92 is 24 more than 68. Ratios can also be calculated, such as stating that 80 is four times 20. In this context, adding, subtracting, multiplying, and dividing the variables is permissible.

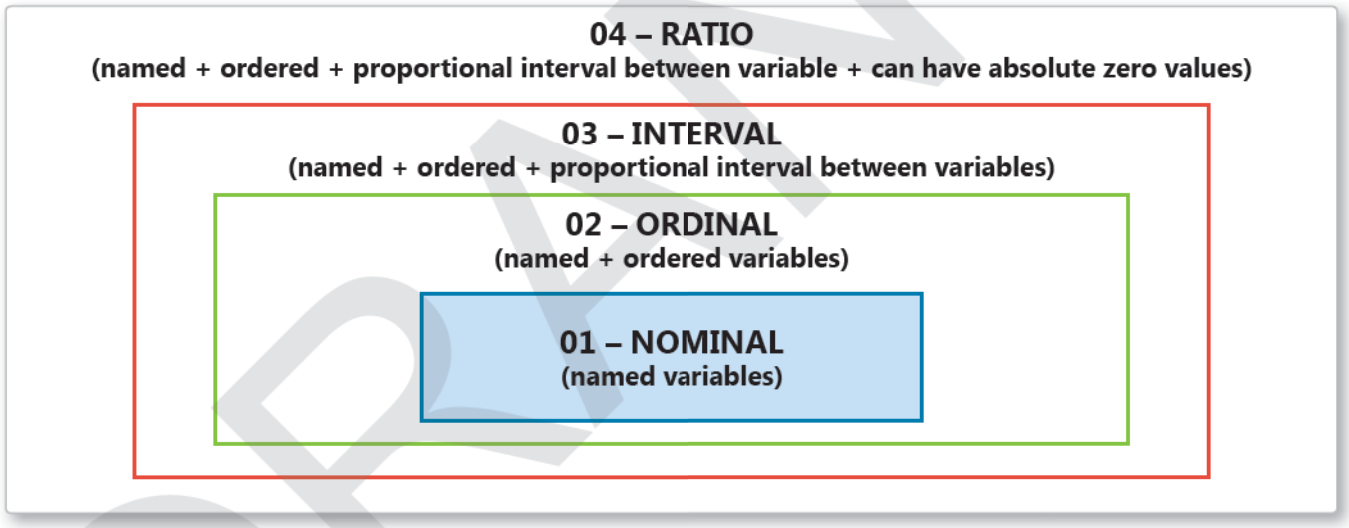
Example:

Range	Frequency
55 – 75	1
76 – 85	1
86 – 95	1
> 95	0

Another example is the weight of a person, where zero weight signifies an absence of weight, that allows meaningful comparisons through addition, subtraction, multiplication, and division.

Comparing All Four Levels of Measurements

Characteristic	Nominal Scale	Ordinal Scale	Interval Scale	Ratio Scale
Type of Data	Categorical	Categorical	Numeric	Numeric
Order/Ranking	No	Yes	Yes	Yes
Equal Intervals	No	No	Yes	Yes
Meaningful Zero Point	No	No	No	Yes
Nature of data	Qualitative	Qualitative or Quantitative	Quantitative	Quantitative
Examples	Gender (Male, Female), Marital Status (Married, Unmarried)	Rank (1st, 2nd, 3rd), Likert Scale	Temperature (Celsius, Fahrenheit), Years	Age, Weight, Height, Income



Digital Literacy

Video Session

Scan the QR code or go to the following link:

Nominal, Ordinal, Interval & Ratio Data: Simple Explanation With Examples -

https://www.youtube.com/watch?v=5Yh-9xdJzAs

Ai Task

Experiential Learning

Divide students in small groups based on categories.

1. Create a list of variables related to your classmates, such as height, eye colour, hair length and weight. Ensure your list covers nominal, ordinal, interval, or ratio level variables. Define clear response categories that are exhaustive and mutually exclusive. For instance, for eye colour, categories could include blue, brown, green, hazel, and other.

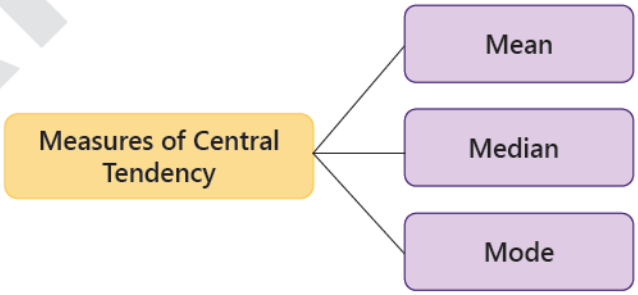
Ai Reboot

Indicate whether the variable is ordinal or not. Write the variable type, if it is not ordinal.

1. Are you eligible to vote?
2. How did you like the session? (1-5)
3. Heights of men (in inches).
4. Grading in exams (A,B,C,D).

Statistical Analysis of Data

Statistical analysis involves collecting, exploring, and presenting large datasets to identify patterns and trends. It's a powerful tool that employs statistical, tests and methodologies to derive meaningful conclusions from raw data. It helps in predicting future trends and making informed decisions based on historical patterns.



Calculating Measure of Central Tendency using Python

Statistical analysis can be performed using the Python programming language. That requires us to import the library statistics into the Python program. Some significant functions that we will utilise in future programs in this module include:

Function	Description
statistics.mean()	Calculates the mean (average) of the provided data.
statistics.median()	Calculates the median (middle value) of the provided data, when arranged in order.
statistics.mode()	Determines the mode of provided numerical or nominal data.
statistics.stdev()	Computes the standard deviation from a sample of data.
statistics.variance()	Calculates the variance of a sample of data.

Mean

The mean in statistics, also known as the average, is a measure of central tendency that represents the sum of all values in a dataset divided by the number of values. It provides a single value that summarises the entire dataset, offering an indication of the typical value.

Mean is an essential concept in statistics. The mean is the average value of a group of numbers.

MEAN (Individual Series): **Direct Method**

Wages (x)
100
200
300
400
500
∑ x = 1500

1. Find ∑x by adding all the values

   ∑x = 1500

2. Find total number of items (N)

   N=5

3. Apply formula

$$\bar{X} = \frac{\sum x}{N} = \frac{1500}{5} = 300$$

MEAN (Discrete Series): (Frequency Array) **Direct Method**

Wages (X)	f	fx
100	2	200
200	1	200
300	3	900
400	3	1200
500	1	500
	$\sum f = 10$	$\sum fx = 3000$

1. Find fx
2. Find $\sum fx$ and $\sum f$
3. Apply formula

$$\bar{X} = \frac{\sum fx}{\sum f}$$

$$= \frac{3000}{10}$$

$$= 300$$

MEAN (Frequency Distribution Series)

Wages (x)	m= $(l_1 + l_2)/2$	f	fm
0 – 100	50	2	100
100 – 200	150	1	150
200 – 300	250	3	750
300 – 400	350	3	1050
400 – 500	450	1	450
		Σf =10	Σfm= 2500

Direct Method

1. Find mid values m= $(l_1 + l_2)/2$
2. Find fm
3. Find Σfm and Σf
4. Apply formula

$$\bar{X} = \frac{\sum fm}{\sum f}$$

$$= \frac{2500}{10}$$

$$= 250$$

Mean Calculation using Python

Program 1: To calculate the mean weight of 25 students.

50.5, 55.2, 60.3, 65.8, 70.1, 75.6, 80.4, 85.7, 90.2, 95.5, 50.3, 55.8, 60.1, 65.4, 70.9, 75.2, 80.6, 85.3, 90.8, 95.1, 50.7, 55.9, 60.5, 65.2, 70.4

Code:

```
import statistics
# List of weights for 25 students
weights = [50.5, 55.2, 60.3, 65.8, 70.1, 75.6, 80.4, 85.7, 90.2, 95.5,
           50.3, 55.8, 60.1, 65.4, 70.9, 75.2, 80.6, 85.3, 90.8, 95.1,
           50.7, 55.9, 60.5, 65.2, 70.4]

# Calculate the mean weight using statistics.mean()
mean_weight = statistics.mean(weights)

# Print the mean weight
print("Mean weight of 25 students is:", mean_weight)
```

Output: Mean weight of 25 students is: 70.46

Median

The **median** is a measure of central tendency that represents the middle value in a dataset when the values are arranged in ascending or descending order. It divides the dataset into two equal halves, with half of the values being less than the median and half being greater.

In statistics, the median is the middle value in a given list of numbers arranged in ascending or descending order.

MEDIAN (Individual Series): with odd number of items

Wages (X)	Wages (X)
500	100
200	200
100	300
300	400
400	500
	n = 5

1. Arrange Data in Ascending Order
2. Median = Size of $\left(\frac{n+1}{2}\right)^{th}$ item

$= \left(\frac{5+1}{2}\right)^{th}$ item

$= 3^{rd}$ item

$= 300$

What if there are even number of items. Lets see!

MEDIAN (Individual Series): with even number of items

Wages (X)	Wages (X)
500	100
200	200
100	300
300	400
600	500
400	600
	n = 6

1. Arrange Data in Ascending Order
2. Median = Size of $\left(\frac{n+1}{2}\right)^{th}$ item

$= \left(\frac{6+1}{2}\right)^{th}$ item

$= 3.5^{th}$ item

$= \frac{3^{rd} \text{ item} + 4^{th} \text{ item}}{2} = \frac{300 + 400}{2}$

$= 350$

MEDIAN: (frequency Distribution Series): (Class Interval Series)

Wages (X)	f	cf
0 – 100	2	2
100 – 200	1	3
200 – 300	3	6
300 – 400	3	9
400 – 500	1	10
	n = 10	

$l_1 = 200$
$cf = 3$
$f = 3$

1. Find Cumulative Frequency (cf)
2. Median Class = Size of $\left(\frac{n}{2}\right)^{th}$ item

   = $(10/2)^{th}$ item = 5^{th} item

   5^{th} item lies in Cumulative Frequency 6
   Hence Median Class = 200 – 300
6. Apply formula

$$M = l_1 + \frac{\frac{n}{2} - cf}{f} \times i$$

$$M = 200 + \frac{\frac{10}{2} - 3}{3} \times 100$$

Median = 266.67

Median calculation using Python

Program 2: To calculate the median weight of 25 students.

50.5, 55.2, 60.3, 65.8, 70.1, 75.6, 80.4, 85.7, 90.2, 95.5, 50.3, 55.8, 60.1, 65.4, 70.9, 75.2, 80.6, 85.3, 90.8, 95.1,50.7, 55.9, 60.5, 65.2, 70.4

Code:

```
import statistics

# List of weights for 25 students with decimal values
weights = [50.5, 55.2, 60.3, 65.8, 70.1, 75.6, 80.4, 85.7, 90.2, 95.5,
           50.3, 55.8, 60.1, 65.4, 70.9, 75.2, 80.6, 85.3, 90.8, 95.1,
           50.7, 55.9, 60.5, 65.2, 70.4]
# Calculate the median weight using statistics.median()
median_weight = statistics.median(weights)
# Print the median weight
print("Median weight of 25 students is:", median_weight)
```

Output: Median weight of 25 students is: 70.1

Mode

The **mode** is a measure of central tendency that identifies the most frequently occurring value in a dataset. Unlike the mean and median, the mode can be used with both numerical and categorical data. A dataset can have one mode (unimodal), more than one mode (bimodal or multimodal), or no mode at all if no number repeats.

In statistics, the mode is the value that appears most often in a given list of numbers.

Inspection Method

MODE (Individual Series):

Marks	Marks	
2	2	Two times
5	2	
7	3	One time
5	5	
6	5	Three times
2	5	
5	6	One time
7	7	
3	7	
7	7	Four times
7	7	

Inspection Method

Arrange the Data in Ascending Order

By Inspection of the Data it is clear that the value '7' occurs most number of times in the series i.e. 4 times

Hence Mode = 7

Mode calculation using Python

Program 3: To calculate the mode of 25 students weight.

50.5, 55.2, 60.3, 65.8, 70.1, 75.6, 80.4, 85.7, 90.2, 95.5, 50.3, 55.8, 60.1, 65.4, 70.9, 75.2, 80.6, 85.3, 90.8, 95.1,50.7, 55.9, 60.5, 65.2, 70.4

Code:

```
import statistics

# New list of weights for 25 students (with mode intentionally set to 60.5)
weights = [50.5, 55.2, 60.3, 65.8, 70.1, 75.6, 80.4, 85.7, 90.2, 95.5,
           50.3, 55.8, 60.5, 65.4, 70.9, 75.2, 80.6, 85.3, 90.8, 95.1,
           50.7, 55.9, 60.5, 65.2, 70.4]

# Calculate the mode weight using statistics.mode()
mode_weight = statistics.mode(weights)

# Print the mode weight
print("Mode weight of 25 students is:", mode_weight)
```

Output: Mode weight of 25 students is: 60.5

When to use mean, median and mode:

Mean	Median	Mode
Data is Normally Distributed: The mean is most informative when the data follows a normal distribution, as it provides a good central point.	**Skewed Data:** The median is better than the mean for skewed distributions because it is not affected by extreme values.	**Categorical Data:** The mode is the only measure that can be used for categorical (nominal) data, indicating the most common category.

Mean	Median	Mode
No Outliers: The mean is sensitive to outliers. If used with the data having no extreme values, could skew the results.	**Outliers Present:** When the dataset contains outliers, the median provides a better central tendency measure.	**Multimodal Distributions:** When the distribution has multiple peaks, the mode helps to identify the frequent values.
Quantitative Analysis: It's useful for further statistical analysis, such as variance and standard deviation calculations.	**Ordinal Data:** When dealing with ordinal data (ranked data), the median is appropriate as it considers the order but not the magnitude of differences.	**Simple Understanding:** It provides a straightforward understanding of the most common value.
Example: • Calculating the average score of students in a class. • Determining the average income of a population when there are no extreme income disparities.	**Example:** • Analysing household income in a region with a few very high incomes. • Reporting the central value in real estate prices, where there can be a significant range.	**Example:** • Identifying the most common brand preference in a survey. • Finding the most frequent diagnosis in a medical dataset.

Variance and Standard Deviation

Measures of central tendency (mean, median, and mode) provide the central value of the data set. Variance and standard deviation are measures of dispersion (quartile, percentile, range). They provide information about the distribution of data around the center.

In this section, we will learn two other measures of dispersion: variance and standard deviation.

Variance

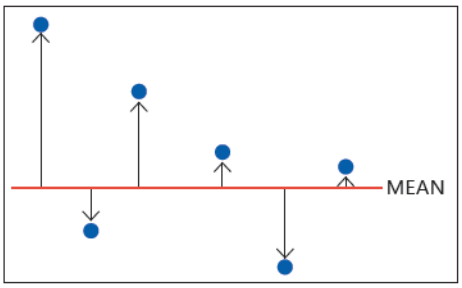
Variance measures the distance of each number in the data set from the mean and also from every other number in the set. Variance is often depicted by the symbol: σ^2.

Calculating the variance:

1 2 3 4 5 6
10, 8, 10, 8, 8, 4
n = 6

10 + 8 + 10 + 8 + 8 + 4 = 48
48 ÷ n = 48 ÷ 6
MEAN = 8

- The variance represents how far the data in your sample are grouped around the mean.
- Data sets with low variance have data grouped closely about the mean.
- Data sets with high variance have data grouped far from the mean.



Step 1: Subtract the mean from each of your numbers in your sample.

10	8	10	8	8	4
−8	−8	−8	−8	−8	−8
2	0	2	0	0	4

Step 2: Square of all the differences.

10	8	10	8	8	4
−8	−8	−8	−8	−8	−8
2^2	0^2	2^2	0^2	0^2	4^2
4	0	4	0	0	16

Step 3: Add all the squared numbers together. This number is called the sum of squares.

10	8	10	8	8	4
−8	−8	−8	−8	−8	−8
2^2	0^2	2^2	0^2	0^2	4^2

4 + 0 + 4 + 0 + 0 + 16

=24

Step 4: Divide the sum of squares by sample size -1 (n-1).

24 ÷ N − 1 = 24 ÷ 5

VARIANCE = 4.8

This value is the variance.

Standard Deviation

Standard deviation is a statistic that measures the degree of dispersion of a data set relative to its average. When determining the deviation of each data point from the average, the standard deviation is calculated as the square root of the variance. If the data points are farther from the mean, the deviation in the data would be greater therefore, the data, would be more spread out, the larger is the standard deviation.

Using the previous example, we have to calculate the square root of the variance in order to find the standard derivation.

VARIANCE = 4.8

STANDARD DEVIATION = $\sqrt{4.8}$ = 2.19

Some key facts concerning variance and standard deviation

- Small variance suggests data points are close to the mean and to each other.
- High variance suggests data points are widely dispersed from the mean and to one another.
- Data points with a low standard deviation are near to the mean, whereas those with a high standard deviation show a wide range of values.

Calculating Variance and Standard Deviation using Python

Program 4: To calculate the variance and standard deviation in the weight of 25 students.

50.5, 55.2, 60.3, 65.8, 70.1, 75.6, 80.4, 85.7, 90.2, 95.5, 50.3, 55.8, 60.1, 65.4, 70.9, 75.2, 80.6, 85.3, 90.8, 95.1,50.7, 55.9, 60.5, 65.2, 70.4

Code:

```
import statistics
# List of weights for 25 students
weights = [50.5, 55.2, 60.3, 65.8, 70.1, 75.6, 80.4, 85.7, 90.2, 95.5,
           50.3, 55.8, 60.1, 65.4, 70.9, 75.2, 80.6, 85.3, 90.8, 95.1,
           50.7, 55.9, 60.5, 65.2, 70.4]
```

```
# Calculate the variance using statistics.variance()
variance_weight = statistics.variance(weights)

# Calculate the standard deviation using statistics.stdev()
std_dev_weight = statistics.stdev(weights)

# Print the variance, and standard deviation
print("Variance of weights:", variance_weight)
print("Standard deviation of weights:", std_dev_weight)
```

Output:

Variance of weights: 207.52666666666664

Standard deviation of weights: 14.405785874663923

Brainy Fact

In machine learning, statistical measurements such as mean, median, and standard deviation are used to analyse data distribution and identify outliers. Data scientists investigate the outliers to see if they are caused by data entry errors, measurement errors, or actual abnormalities, and then decide whether to remove, correct, or maintain them based on their significance to the analysis.

AI Reboot

Fill in the blanks:

1. Mean, median and mode provide the _____ value of the dataset while variance and standard deviation provide information about the _____ of data around the centre.
2. It is better to use the _____ in multimodal distributions.
3. The mean in Python can be calculated using _____ function.
4. Datasets with _____ variance have data grouped closely about the mean.
5. The mean in statistics is also known as the _____.

Representation of Data

Statistics is a branch of mathematics that involves the collection, analysis, interpretation, presentation, and organisation of data. It is used to make informed decisions and understand the world through data. To accomplish this goal, statisticians summarise a significant amount of data in a compact format that yields relevant results. Without displaying values for each observation (from populations), it is possible to portray the data in a concise manner while retaining its significance using techniques known as 'data representation'. It may also be defined as a technique for presenting enormous amounts of data in a way that allows the user to quickly and easily interpret the most relevant information.

There are two broad categories of data representation techniques:

- **Non-Graphical Technique:** Non-graphical techniques include tabular and case forms. This is an older data representation format that is unsuitable for huge datasets. Non-graphical strategies are ineffective when we want to make decisions based on a set of data.

- **Graphical Technique:** Graphs are commonly used to visualise statistical data using points, lines, dots, and other geometric shapes. The human brain is more comfortable coping with complex and massive amounts of material when it is represented visually. **Data visualisation** refers to the graphical or pictorial depiction of data using graphs, charts, and other tools.

Ai Task

Creativity and Innovativeness

Visit https://datavizcatalogue.com/ and study the different types of charts available.

The Data Visualisation Catalogue is a project developed by Severino Ribecca to create a (non-code-based) library of different information visualisation types. The website serves as a learning and inspiration resource for those working with data visualisation.

Introduction to Matplotlib

Visualisation in Python can be accomplished using the Matplotlib library. This extensive library enables the creation of various plots, such as line plots, bar charts, histograms, scatter plots, and more. Matplotlib is highly customisable, giving users detailed control over the appearance of the plots. The 'pyplot,' submodule of Matplotlib, offers a MATLAB-like interface and includes numerous convenience functions that simplify the process of creating basic plots.

Install Matplotlib library in Python by giving the following commands:

```
pip install matplotlib
or
python - m pip install - U matplotlib
```

While writing a Python program, we import the pyplot module of matplotlib library by giving the following command:

```
import matplotlib.pyplot
```

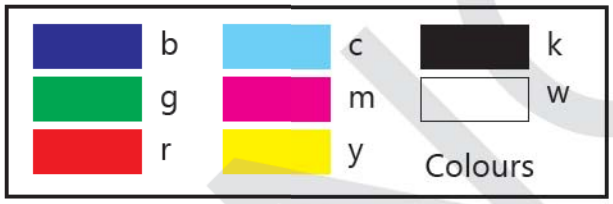
Some of the common functions of Matplotlib library with their descriptions is given below:

Function Name	Description
`title( )`	Adds title to the chart/graph
`xlabel( )`	Gives label for X-axis
`ylabel( )`	Gives label for Y-axis
`xlim( )`	Gives the value limit for X-axis
`ylim( )`	Gives the value limit for Y-axis
`xticks( )`	Places the tick marks on the X-axis
`yticks( )`	Places the tick marks on the Y-axis
`show( )`	Displays the graph on the screen
`savefig("address")`	Saves the graph to the given address
`figure(figsize = value in tuple format)`	Sets the size of the plot where the graph is drawn. Values should be supplied in tuple format to the figsize attribute, which is passed as an argument.

Markers in pyplot

Marker	Symbol	Description
"."	●	point
","	.	pixel
"o"	●	circle
"v"	▼	triangle_down
"^"	▲	triangle_up
"s"	■	square
"p"	⬟	pentagon

Marker	Symbol	Description
"p"	✚	plus (filled)
"*"	★	star
"h"	⬢	hexagon1
"H"	⬢	hexagon2
"+"	+	plus
"x"	×	x
"0"	◆	diamond



Experiential Learning

Video Session

Watch the following video: Python Tutorials - Making a Simple Plot Using pyplot module | matplotlib-
https://www.youtube.com/watch?v=flwF6aJtmJs&list=PLzgPDYo_3xulakyk7r5h_djrWq1gjD6hm&index=2

Create the plot explained in the video.

Brainy Fact

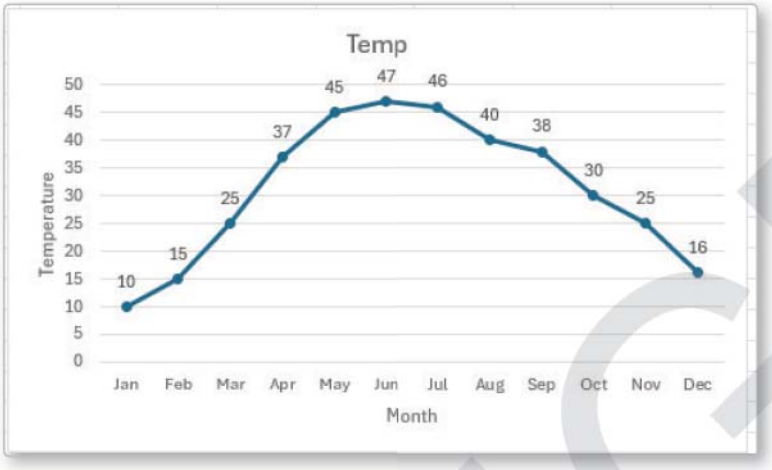
Visualisations are often the simplest way to convey facts! Psychologically, our brain experiences less stress when viewing graphical representations compared to examining the same data in a numerical list format.

Line Graph

A line graph is a strong tool for representing **continuous data** on a numbered axis. It enables us to visually represent **trends and changes** in data points across time. Line graphs are appropriate for data that can take any value within a defined range. The line may slope upwards, suggesting an increase, or downwards, indicating a decrease, reflecting changes in the data over time.

Example: Let us consider the average temperatures in Delhi and then draw a line graph from the available data:

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperature	10	15	25	37	45	47	46	40	38	30	25	16



Program 5: Create a Line Graph for temperatures in different months using Python.

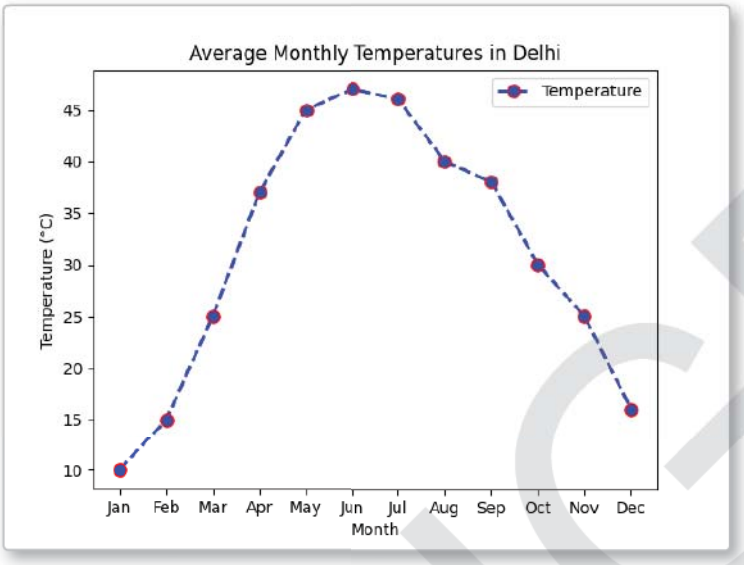
Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperature	10	15	25	37	45	47	46	40	38	30	25	16

Code:

```
import matplotlib.pyplot as plt

# Sample data for the line graph
months = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov',
'Dec']
temperatures = [10, 15, 25, 37, 45, 47, 46, 40, 38, 30, 25, 16]

# Create a line graph using plot() function with customized markers, linewidth, and
linestyle
plt.plot(months, temperatures,
         marker='o', markersize=8, markeredgecolor='red',
         linestyle='--', linewidth=2, color='b', label='Temperature')

# Add titles and labels
plt.title('Average Monthly Temperatures in Delhi')
plt.xlabel('Month')
plt.ylabel('Temperature (°C)')

# Show the legend
plt.legend()
```

```
# Display the plot
plt.show()
```

Output:



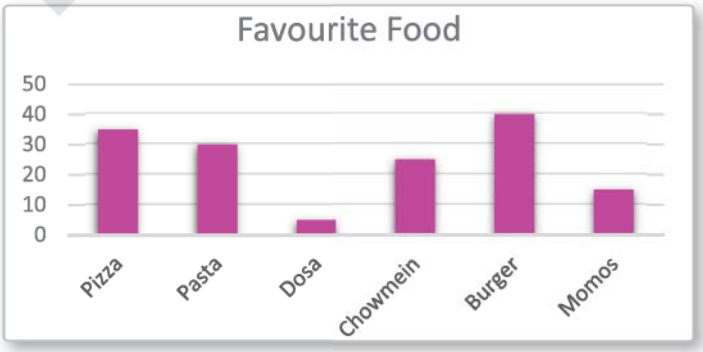
In the above program the following attributes have been used:

- linewidth - the width of the line
- linestyle - the style of line i.e. solid, dashed, dot, dashdot marker, markersize, and markeredgecolor - sets the marker's shape, size and marker edge colour, respectively
- marker - sets the marker style for the data points
- markersize - sets the size of the markers
- color - sets the colour of the line
- label - sets the label for the line plot

Bar Graph

A bar chart, sometimes referred to as a horizontal column chart, is well-liked for a reason that it swiftly visualises data sets and is easy on the eyes to visualise. You can rapidly determine which bar is the highest or lowest as well as the incremental changes between bars. The bar chart is best suited to represent categorical data. For example:

Food	Pizza	Pasta	Dosa	Chowmein	Burger	Momos
People	35	30	5	25	40	15



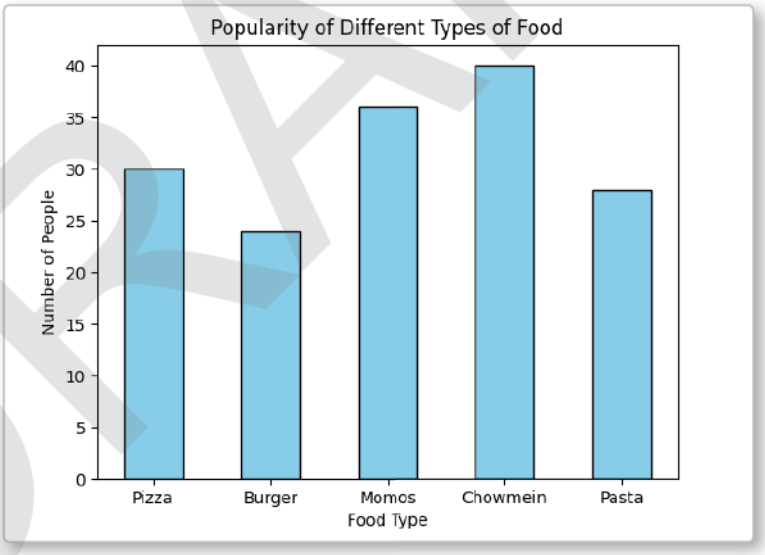
Program 6: Create a Bar Graph for how many people like varied food items using Python.

Food	Pizza	Burger	Momos	Chowmein	Pasta
No. of People	30	24	36	40	28

Code:

```
import matplotlib.pyplot as plt

# Sample data for the bar chart
food_types = ['Pizza', 'Burger', 'Momos', 'Chowmein', 'Pasta']
values = [30, 24, 36, 40, 28]

# Create a bar chart using bar() function with customized width
plt.bar(food_types, values, color='skyblue', edgecolor='black', width=0.5)

# Add titles and labels
plt.title('Popularity of Different Types of Food')
plt.xlabel('Food Type')
plt.ylabel('Number of People')

# Display the plot
plt.show()
```

Output:



The various attributes used in the above program are:

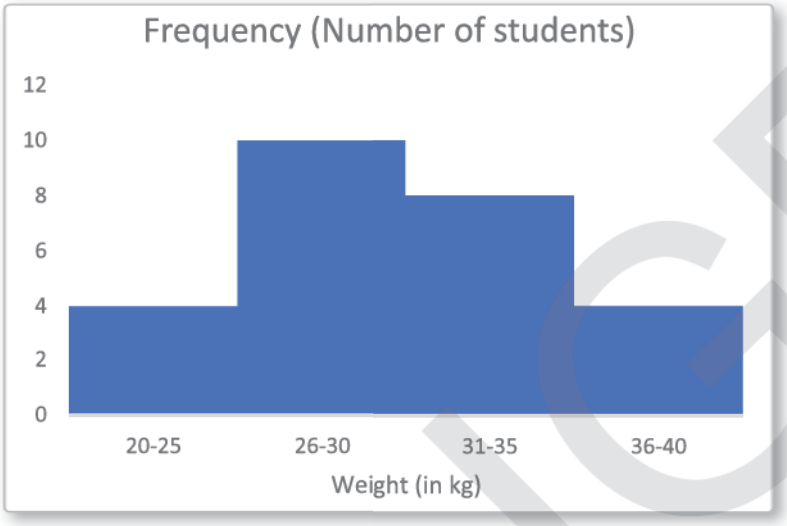
- color – sets the bar colour
- edgecolor – sets the colour of the bar edges
- width – sets the width of the bars

Histogram

Histogram is a form of bar graph. It is a visualisation of several outcomes organised into columns along the x-axis. The y-axis of the histogram represents the number count or multiple occurrences in the data for each column.

Histogram is the simplest method for visualising data distributions. For example:

Weights (in kg)	Frequency (Number of students)
20-25	4
26-30	10
31-35	8
36-40	4



Program 7: Create a Histogram for number of hours students spend studying per week using Python Number of hours students spend studying per week:

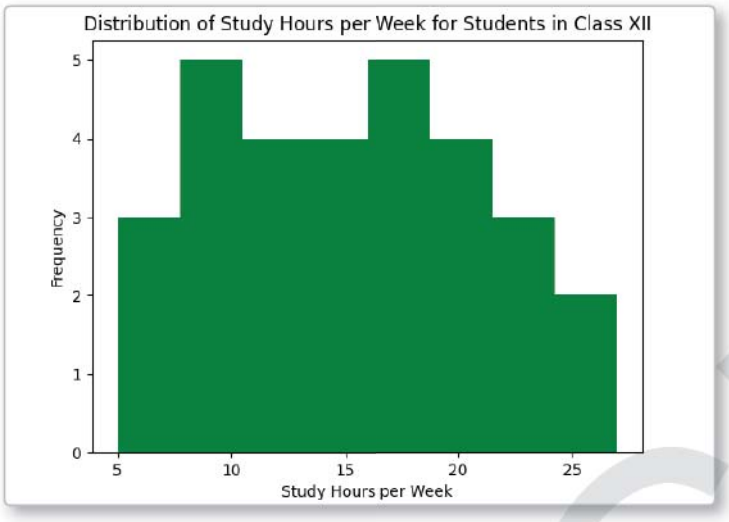
[5, 7, 8, 10, 12, 14, 15, 16, 18, 20, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24]

This example takes number of hours per week that students spend in studying. To create a histogram from the given data, we must first organise it into intervals. These intervals are frequently referred to as logical ranges or bins.

Code:

```
import matplotlib.pyplot as plt

# Sample data: number of hours students spend studying per week
study_hours = [5, 7, 8, 10, 12, 14, 15, 16, 18, 20,
               9, 11, 13, 15, 17, 19, 21, 23, 25, 27,
               6, 8, 10, 12, 14, 16, 18, 20, 22, 24]

# Create a histogram using hist() function
plt.hist(study_hours, bins=8, color='green')

# Add titles and labels
plt.title('Distribution of Study Hours per Week for Students in Class XII')
plt.xlabel('Study Hours per Week')
plt.ylabel('Frequency')

# Display the plot
plt.show()
```

Output:

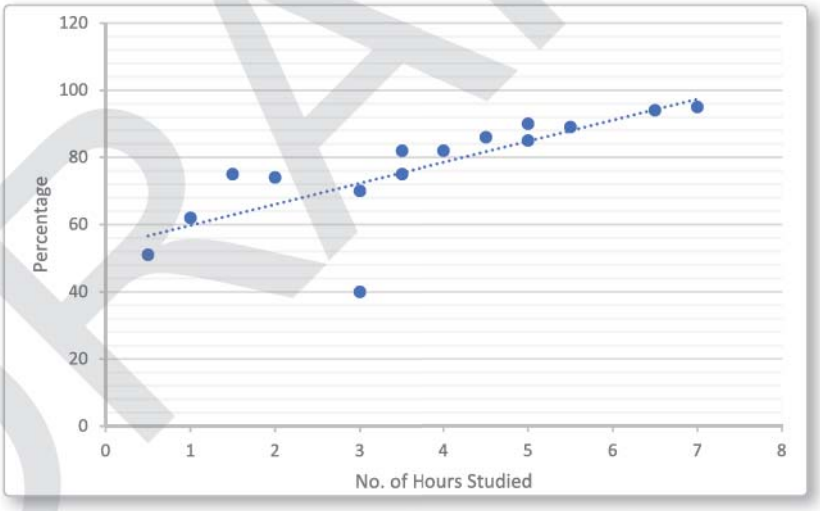


The choice of bins is subjective and can be adjusted based on specific data and visualisation needs. You can experiment the above example with bins=5 or bins=10 also.

Scatterplot

The relationship between items depending on two separate variables and data sets can be shown using a scatter plot or scatter chart. An x-y coordinate system is used to plot dots (or plot data). A trend line may be added to a scatter plot in some instances. For example:

Study Time (hours)	4	3.5	5	2	3	6.5	0.5	3.5	4.5	5	1	1.5	3	5.5	7
Percentage	82	82	90	74	40	94	51	75	86	85	62	75	70	89	95



The most popular data visualisation method in machine learning is the scatter plot.

When is a scatter plot suitable?

- It is used to track the connections between two numerical variables. When the data is seen as a whole, the dots on the plot indicate both the variable's value and any trends.
- The scatterplot is a valuable tool for calculating correlation. Variable relationships can be categorised in a variety of ways, including positive or negative, strong or weak, linear or nonlinear.
- This graph's strength lies in its ability to clearly depict trends, clusters, and relationships within datasets.

The function scatter() is used to visualise a scatterplot in Python.

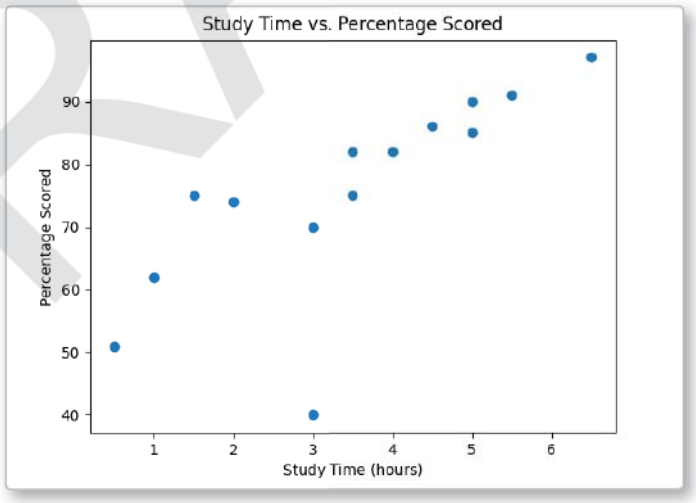
Program 8: Create a Scatter Plot Graph for number of hours students spend studying per week with respect to percentage scored using Python

Study Time	4	3.5	5	2	3	6.5	0.5	3.5	4.5	5	1	1.5	3	5.5
percentage	82	82	90	74	40	97	51	75	86	85	62	75	70	91

Code:

```
import matplotlib.pyplot as plt

# Data
study_time = [4, 3.5, 5, 2, 3, 6.5, 0.5, 3.5, 4.5, 5, 1, 1.5, 3, 5.5]
percentage = [82, 82, 90, 74, 40, 97, 51, 75, 86, 85, 62, 75, 70, 91]

# Create scatter plot using scatter() function
plt.scatter(study_time, percentage)

# Add title and labels
plt.title('Study Time vs. Percentage Scored')
plt.xlabel('Study Time (hours)')
plt.ylabel('Percentage Scored')

# Show plot
plt.show()
```

Output:



Many times, data is provided through CSV files. So, let us learn to visualise a scatterplot by uploading the CSV file.

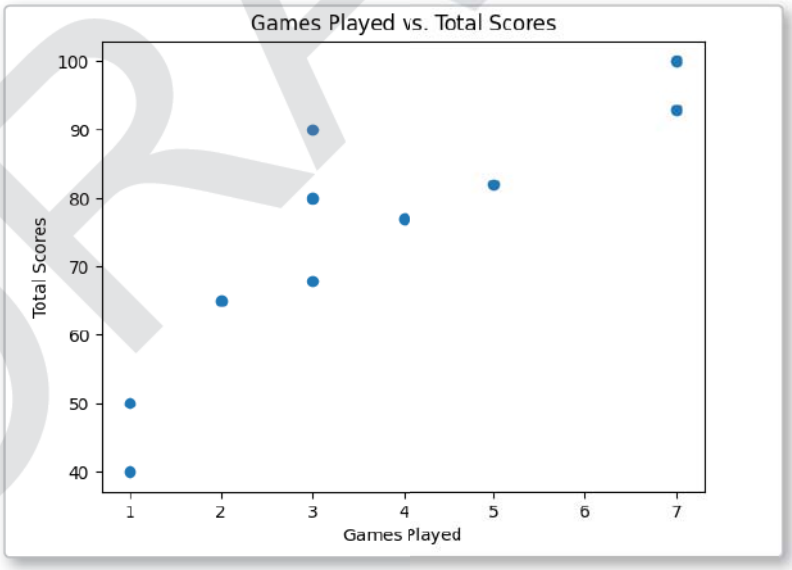
Program 9: Create a Scatter Plot Graph from CSV file saved as "Games.csv' for games played and scored attained using Python. Contents of 'Games.csv' file are as follows:

Games Played	**5**	**3**	**4**	**2**	**7**	**1**	**3**	**1**	**7**	**3**
Total Scores	82	90	77	65	93	50	68	40	100	80

Code:

```
import matplotlib.pyplot as plt
import pandas as pd

# Load the CSV file
data = pd.read_csv('Games.csv')

# Extract the data for the scatter plot
games_played = data['Games Played']
total_scores = data['Total Scores']

# Create scatter plot using scatter() function
plt.scatter(games_played, total_scores)

# Add title and labels
plt.title('Games Played vs. Total Scores')
plt.xlabel('Games Played')
plt.ylabel('Total Scores')

# Show plot
plt.show()
```

Output:



Let us understand this program:

- We use the pandas library function 'read' to read a CSV file named Games.csv and load its contents into a **DataFrame** called data. A DataFrame is a two-dimensional, size-mutable, and potentially heterogeneous tabular data structure with labeled axis (rows and columns).

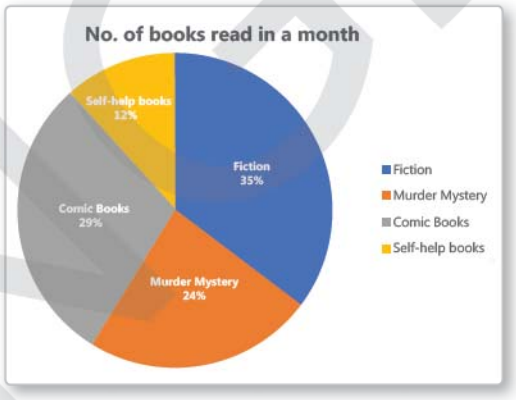
- 'games_played' extracts the column named 'Games Played' from the DataFrame, creating a series that contains the number of games played.
- 'total_scores' extracts the column named 'Total Scores' from the DataFrame, creating a series that contains the total scores.
- These extracted series (games_played and total_scores) are then used to create a scatter plot.

Pie Chart

A pie chart is a circular chart with the pie divided into many segments or sections. Each division of the pie indicates the relative size, that is, the contribution of every class or a percent of the total. Pie charts are mostly used to visualise facts from a small dataset. There should be not more than seven categories in a pie chart. Another limitation of a Pie chart is that zero values can't be displayed in pie chart. However, these graphs are hard to interpret and compare facts with another pie chart.

Pie charts are used to examine elements of a whole. They do not display changes over time. For example, pie charts may be used to suggest the achievement or failure of a product or service, display the time-period allotted to each subject in a class, or depict monthly spending on numerous items and services in a household. For example:

Story books	No. of books read in a month
Fiction	30
Murder Mystery	20
Comic Books	25
Self-help books	10



The function pie() is used to plot a pie chart.

Program 10: Create a Pie chart for different category of books read in a month using Python.

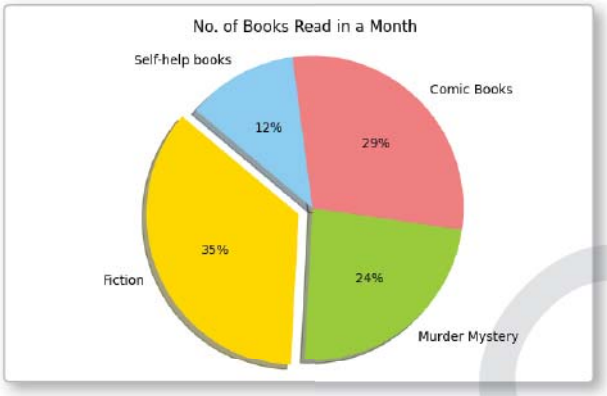
Books Category	Fiction	Murder Mystery	Comic Books	Self-help books
Read in a Month	30	20	25	10

Code:

```
import matplotlib.pyplot as plt

# Data
labels = ['Fiction', 'Murder Mystery', 'Comic Books', 'Self-help books']
sizes = [30, 20, 25, 10]
colors = ['gold', 'yellowgreen', 'lightcoral', 'lightskyblue']
explode = (0.1, 0, 0, 0)  # explode the first slice (Fiction)

# Create pie chart using pie() function.
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
        autopct='%1.0f%%', shadow=True, startangle=140)

# Add title
plt.title('No. of Books Read in a Month')
```

```
# Equal aspect ratio ensures that pie is drawn as a circle.
plt.axis('equal')

# Show plot
plt.show()
```

Output:



This code will generate a pie chart with slices representing the number of books read in a month for each category: Fiction, Murder Mystery, Comic Books, and Self-help books. The first slice (Fiction) is exploded to highlight it explode parameter. The 'autopct' parameter displays the percentage value of each slice (each label will show the percentage value of the corresponding slice, rounded to zero decimal place), and 'shadow' adds a shadow effect. The 'startangle' parameter rotates the start of the pie chart to 140 degrees for better visual appeal.

Ai Task

Experiential Learning

1. Learn to create a pie chart of the 3 species of the Iris Dataset. Watch the video, How to visualise Pie-chart using Python | Pie chart tutorial - https://www.youtube.com/watch?v=9cFsCfKLyrQ
   Now create the pie chart on your own.
2. Next take a quiz on Matplotlib. Scan the QR code or go to the website https://runestone.academy/ns/books/published/py4e-int/viz/Exercises.html
   How much did you score?

Introduction to Matrices

The knowledge of matrices is necessary in all branches of Mathematics. Matrix is one of the most powerful tools in Mathematics. Compared to other methods, it greatly simplifies the calculations.

In mathematics, a matrix (plural matrices) is a rectangular arrangement of numbers. The numbers are arranged in tabular form as rows and columns.

Let us understand with the help of an example:
Ganga bought 12 oranges and 5 apples.
Hemant bought 10 oranges and 7 apples.
Gargi bought 5 oranges and 2 apples.

The above information can be arranged in tabular form as follows:

	Oranges	Apples
Ganga	12	5
Hemant	10	7
Gargi	5	2

And this can be represented as:

$$\begin{matrix} \text{Row1} \\ \text{Row2} \\ \text{Row3} \end{matrix}\begin{bmatrix} 12 & 5 \\ 10 & 7 \\ 5 & 2 \end{bmatrix}$$
$$\text{Col1} \quad \text{Col2}$$

The entries in the rows represent number of oranges and apples bought by Ganga, Hemant and Gargi respectively. It can also be represented in another form as:

$$\begin{bmatrix} 12 & 10 & 5 \\ 5 & 7 & 2 \end{bmatrix} \begin{matrix} -\text{Row1} \\ -\text{Row2} \end{matrix}$$
$$\text{Col1} \quad \text{Col2} \quad \text{Col3}$$

Here, the entries in the columns represent the number of oranges and apples bought by Ganga, Hemant and Gargi respectively.

We denote matrices by capital letters. For example:

$$A = \begin{bmatrix} 2 & 7 \\ 0 & \sqrt{5} \\ -3 & 1 \end{bmatrix}$$

Order of Matrix

A matrix having m rows and n columns is called a matrix of the *order* $m \times n$ or simply $m \times n$ matrix (read as an m by n matrix). So, referring to the above example, matrix A is a 3×2 matrix. The number of elements present in a matrix is given by $m \times n \Rightarrow 3 \times 2 = 6$ elements.

Each individual element is represented as a_{ij} where i represents row and j represents column.

$$A = \begin{bmatrix} 2_{a_{11}} & 4_{a_{12}} & 3_{a_{13}} \\ 7_{a_{21}} & 5_{a_{22}} & 9_{a_{23}} \end{bmatrix}$$

In general, a_{ij}, is an element lying in the i^{th} row and j^{th} column. We can also call it as the $(i, j)^{th}$ element of the matrix.

Ai Task

Problem Solving & Logical Reasoning

Given the following matrix, identify the following elements:

$$A = \begin{bmatrix} 7 & 6 & -5 \\ 4 & 0 & 2 \\ -3 & 8 & 10 \end{bmatrix}$$

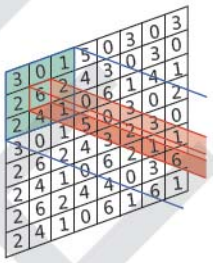
a. a_{23} b. a_{11} c. a_{31}

Brainy Fact

We all know that computer understands data only in the form of binary/hexadecimal numbers. Then, how are images processed by computers?

Consider, the given image:

On the computer/mobile, the image is represented as a combination of pixels. This is represented mathematically as matrices!



Operations on Matrices

Addition of Matrices

The sum of two matrices is a matrix obtained by adding the corresponding elements of the given matrices. Also, the two matrices have to be of the same order. For example:

$$A=\begin{bmatrix}2 & 4\\ 6 & 8\\ 7 & 5\end{bmatrix}_{3\times 2} \quad B=\begin{bmatrix}1 & 0\\ -2 & 7\\ 3 & 4\end{bmatrix}_{3\times 2}$$

$$A+B=\begin{bmatrix}2+1 & 4+0\\ 6+(-2) & 8+7\\ 7+3 & 5+4\end{bmatrix}\Rightarrow\begin{bmatrix}3 & 4\\ 4 & 15\\ 10 & 9\end{bmatrix}$$

Difference of Matrices

If A and B are two matrices of the same order, then the difference A – B is defined as a matrix where each element is obtained by subtracting the corresponding elements $(a_{ij} - b_{ij})$. For example:

$$A=\begin{bmatrix}1 & 2 & 3\\ 2 & 3 & 0\end{bmatrix}_{2\times 3} \quad B=\begin{bmatrix}3 & -2 & 3\\ -1 & 0 & -2\end{bmatrix}_{2\times 3}$$

$$A-B=\begin{bmatrix}1-3 & 2-(-2) & 3-3\\ 2-(-1) & 3-0 & 0-(-2)\end{bmatrix}\Rightarrow\begin{bmatrix}-2 & 4 & 0\\ 3 & 3 & 2\end{bmatrix}_{2\times 3}$$

Transpose of a Matrix

A matrix obtained by interchanging the rows and columns of a matrix. Transpose of a matrix A is denoted by A′ or A^T. If order of matrix A is m X n, then order of transpose of matrix A, i.e., A′ is n x m. For example:

$$A=\begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{bmatrix}_{2\times 3} \quad A^T \text{ OR } A'=\begin{bmatrix}1 & 4\\ 2 & 4\\ 3 & 6\end{bmatrix}_{3\times 2}$$

Multiplication of a Matrix by a Scalar

A scalar is any number. So, if A is a matrix and *k* is a scalar, then *k*A is another matrix which is obtained by multiplying each element of A by the scalar *k*.

For example:

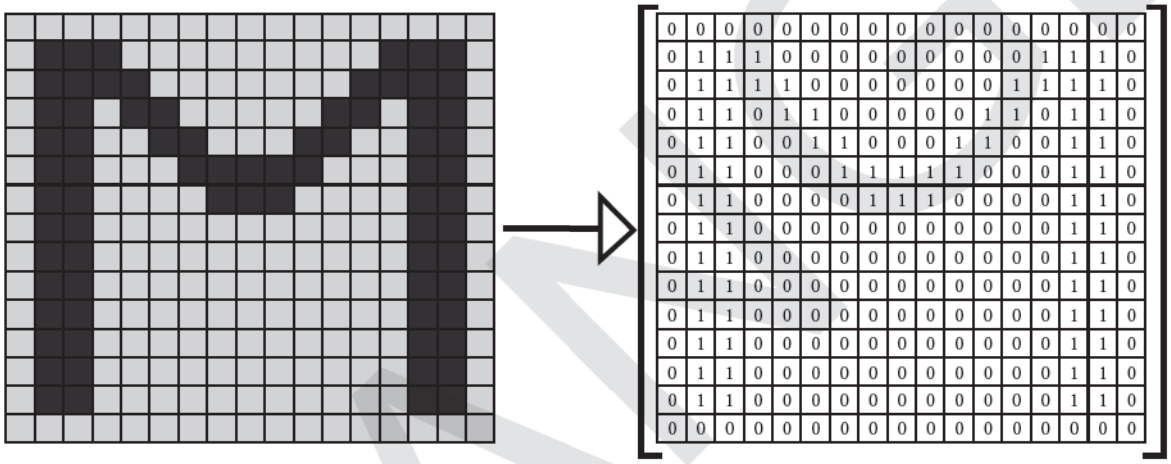
$$A = \begin{bmatrix} 1 & 5 & 3 \\ 2 & 4 & 7 \end{bmatrix} k = 3$$

$$kA = 3A = \begin{bmatrix} 1\times3 & 5\times3 & 3\times3 \\ 2\times3 & 4\times3 & 7\times3 \end{bmatrix} \Rightarrow \begin{bmatrix} 3 & 15 & 9 \\ 6 & 12 & 21 \end{bmatrix}$$

Applications of Matrices in AI

Matrices are used throughout the field of machine learning for computing:

- **Sales Forecasting and Price Prediction**: Matrices are used to represent relevant predictor and response variables.
- **Image Processing**: Digital images can be represented using matrices. In the following figure, each box appears transparent or coloured. It is represented in the corresponding matrix with either a "1" or a "0". A zero means the box contains nothing and is white, while a one indicates that the box is filled (in this case it is filled with black).



- **Recommender Systems:** They use matrices to relate between users and the purchased or viewed item(s).
- **Natural Language Processing**: In NLP, vectors represent the distribution of a particular word in a document. Vectors are one-dimensional matrices.

Data Preprocessing

Data preprocessing is an essential phase in the machine learning process that prepares datasets for effective machine learning applications. It includes multiple processes to clean, transform, reduce, integrate, and normalise data:

1. **Data Cleaning:** Businesses have abundance of data. However, not all of it is accurate or organised. When it comes to machine learning, if data is not sufficiently cleaned, the accuracy of your model is at risk. The following steps are taken to clean/prepare the data:
   - **Missing Data:** Missing data refers to the absence of certain values in the dataset, which can result from various causes. To handle missing data, strategies include removing rows or columns with missing values, imputing missing values with estimates, or utilising algorithms that can manage missing data.

- **Outliers (extreme values):** Outliers are data points that deviate significantly from most of the dataset, typically due to errors or uncommon occurrences. Managing outliers includes detecting and excluding them, transforming the data, or applying robust statistical techniques to minimise their influence.
- **Inconsistent Data:** Inconsistent data, such as typographical errors or variations in data types, is rectified to ensure uniformity and coherence across the dataset.
- **Duplicate Data:** Duplicate data is identified and eliminated to maintain data integrity and accuracy.

2. **Data Transformation:** This process involves converting data into a format suitable for analysis. Common techniques include normalisation, standardisation, and discretisation. Normalisation scales the data to a common range, standardisation adjusts the data to have a zero mean and unit variance, and discretisation converts continuous data into discrete categories. Existing features may also be adjusted as necessary.

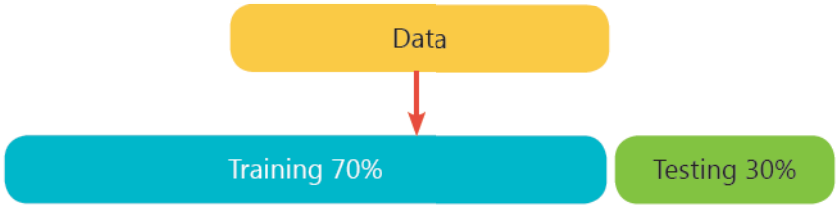
3. **Data Reduction:** This process decreases the data volume, making analysis easier while yielding the same or nearly the same results. It also helps to save storage space. Common data reduction techniques include dimensionality reduction (reducing the number of features in a dataset) and data compression.

4. **Data Integration and Normalisation:** Data from multiple sources or formats is combined or aggregated (data is presented in the form of a summary). Subsequently, the data is normalised to ensure uniform scale and distribution across all features, enhancing the effectiveness of machine learning models. Data integration is a key component of data management.

5. **Feature Selection:** This step involves choosing a subset of important features from the dataset. Feature selection is commonly done to eliminate irrelevant or redundant features from the dataset.

Data in Modelling and Evaluation

Once data preprocessing is complete, it's divided into two sets: the **Training data** and the **Testing data**.



The **Training data** is utilised to teach machine learning models, while the **Testing data** assesses how well the trained models perform. During modelling, suitable machine learning algorithms are selected based on the problem type (e.g., classification, regression, clustering) and dataset characteristics.

Training data vs. Testing data

Feature	Training data	Testing data
Purpose	Training data is a learning phase. The more training data the model has, the better it can make predictions.	Testing Data is used to check the performance of the model.
Exposure	The model learns from the Training data to make accurate predictions.	The Testing data is not exposed to the model before evaluation. Testing data is the new data.
Distribution	The distribution of the Training data should be like the distribution of the real-world data that the model will be used in.	The distribution of the Testing data may be entirely different from the real-world data.
Size	The Training data is larger in size as the model needs to analyse and observe the patterns for making accurate predictions.	The size of the Testing data is smaller than the training data because it is used to evaluate the performance of the model that has been trained on the training data.

Various techniques like train-test split, cross-validation, and error analysis are employed to gauge the model's generalisation ability and pinpoint areas for enhancement. In the **Train Test Split technique**, dataset is divided into two sets: Training and Testing. It trains the model with the Training data and assesses its performance using the Testing data. Cross Validation ensures consistent model performance across different data subsets. You will study these in detail in class XII.

Different evaluation techniques are applied depending on the data type. For classification problems, metrics such as accuracy, precision, recall, F1-score, and ROC curve are commonly used. For regression tasks, metrics like mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared are frequently used.

In today's era, having proficiency in handling data is crucial. With the rise of artificial intelligence, understanding data allows us to leverage information effectively. It's akin to have a map for navigating a large city; being adept with data empowers us to make informed decisions and utilse technology wisely.

At a Glance

- Data literacy involves the ability to find and use data proficiently.
- Data can be structured, semi-structured, or unstructured.
- AI data analysis employs AI techniques and data science to enhance the processes of cleaning, inspecting, and modelling over both structured and unstructured data.
- Data collection means gathering data from many sources, both offline and online.
- Primary and secondary are the two main sources from which data is collected.
- Primary data is obtained directly from the source and has not been previously published or analysed by others.
- Secondary data can be obtained from research articles, books, reports, and internet databases.
- The method used to measure a collection of data is known as the level of measurement.

- Statistical analysis is the process of collecting, exploring, and presenting huge volumes of data in order to identify patterns and trends.
- The term "central tendency" refers to a single number that summarises the complete distribution of a data domain (or data set).
- The mean, the most commonly used measure of central tendency, is the average value of a collection of data points.
- The median is the middle value of a set of data, obtained by ranking all of the data points and selecting the one in the centre.
- Mode is used to find the distribution peak, and there can be multiple peaks.
- Data points with a low standard deviation are near to the mean, whereas those with a high standard deviation show a wide range of values.
- Data representation is defined as a technique for presenting enormous amounts of data in a way that allows user to quickly and easily interpret the most relevant information.
- There are two broad categories of data representation techniques - Non-Graphical Technique and Graphical Technique.
- Data visualisation in Python can be accomplished using the Matplotlib library.
- The 'pyplot' submodule of Matplotlib offers a MATLAB-like interface and includes numerous convenience functions that simplify the process of creating basic plots.
- Data preprocessing is an essential phase in the machine learning process that prepares datasets for effective machine learning applications. It includes multiple processes to clean, transform, reduce, integrate, and normalise data.
- Once data preprocessing is complete, the dataset is divided into two sets: the Training dataset and the Testing dataset.
- The Training dataset is utilised to teach machine learning models, while the Testing dataset assesses how well the trained models perform.
- In today's era, having proficiency in handling data is crucial.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. Data can be described as a representation of _____ .
   a. Random information ◯ b. Facts or instructions about entities ◯
   c. Irrelevant details ◯ d. Only numerical values ◯
2. Which of the following is NOT a primary data source?
   a. Surveys ◯ b. Interviews ◯
   c. Observations ◯ d. Published reports ◯

3. Which data collection method is used to engage in direct connection with individuals or organisations and can be structured, semi-structured, or unstructured?

a. Questionnaire ◯ b. Marketing Campaign ◯

c. Interview ◯ d. Observation ◯

4. Testing the efficiency of different advertising techniques on a group of people is an example of which primary data source?

a. Observation ◯ b. Experiment ◯

c. Marketing Campaign ◯ d. Questionnaire ◯

5. What method employs automated software to gather specific information and data from websites?

a. Social Media Data Tracking ◯ b. Web Scraping ◯

c. Satellite Data Tracking ◯ d. Online Data Platforms ◯

6. Kaggle, GitHub, KDNuggets, and Google Dataset Search are examples of what type of secondary data source?

a. Books, textbooks, and encyclopedias ◯ b. Web Scraping ◯

c. Satellite Data Tracking ◯ d. Online Data Platforms ◯

7. Which level of measurement is used when data is categorised without any order or ranking?

a. Ordinal ◯ b. Interval ◯

c. Ratio ◯ d. Nominal ◯

8. Which measure of central tendency is most affected by extreme values (outliers)?

a. Mean ◯ b. Median ◯

c. Mode ◯ d. All are equally affected ◯

9. When is it most appropriate to use the median as a measure of central tendency?

a. When the dataset contains no outliers ◯

b. When the dataset is skewed or contains outliers ◯

c. When the data is nominal ◯

d. When calculating the average of categorical data ◯

10. What is the output of statistics.stdev ([5, 10, 15, 20, 25])?

a. The average of the data ◯ b. The middle value of the data ◯

c. The most frequent value in the data ◯ d. The standard deviation of the data ◯

11. In a dataset with no repeating values, what can be said about the mode?

a. There is one mode ◯ b. There are multiple modes ◯

c. There is no mode ◯ d. There is a mode, but it cannot be calculated ◯

12. What is the purpose of data visualisation?

a. To make data analysis more complicated ◯ b. To simplify the interpretation of complex data ◯

c. To hide information from the user ◯ d. To increase the size of datasets ◯

13. Which function is used to set the size of the plot in Matplotlib?

a. title() ◯ b. xlabel() ◯

c. figure() ◯ d. show() ◯

14. What does each sector (slice) in a pie chart represent?

a. Frequency or count ◯ b. Mean value ◯

c. Individual data point ◯ d. Percentage of the whole ◯

15. Which data is used to evaluate the performance of the AI model?

a. Training data ◯ b. Test Data ◯

c. Raw data ◯ d. unstructured data ◯

B. Fill in the blanks.

1. The _____ library needs to be imported in Python in order to plot graphs.
2. Data collection can be classified into two categories, namely _____ data, and _____ data.
3. _____ is collecting data on the Earth's surface and atmosphere through satellite.
4. _____ involves collecting, exploring, and presenting large datasets to identify patterns and trends.
5. High variance suggests data points are widely dispersed from the mean and _____ .
6. Training data is _____ in size than testing data.
7. A _____ is a two-dimensional, size-mutable, and potentially heterogeneous tabular data structure with labeled axis (rows and columns).
8. _____ scales the data to a common range.

C. State whether the following statement is True or False.

1. It is a well-known fact that Artificial Intelligence (AI) is fundamentally driven by algorithms. _____
2. Collecting large amounts of data can be the hardest part of a machine learning project. _____
3. Primary data sources might save time and resources. _____
4. Ordinal data consists of categories arranged in a random manner. _____
5. Central tendency helps in predicting future trends and making informed decisions based on historical patterns. _____
6. We can have multiple mode values. _____
7. Non-graphical strategies are ineffective when we want to make decisions based on a set of data. _____
8. xlabel( ) gives the value limit for Y-axis. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. Define a primary data source.

Ans. A primary data source refers to the original source from which data is collected firsthand. This data is obtained directly from its origin, without any intermediary sources or interpretations. Primary data sources include surveys, interviews, observations, experiments, and any other method where data is collected directly by the researcher or organisation for a specific purpose. This type of data is considered valuable because it is tailored to the specific research or business needs and is often more accurate.

2. How does statistical methods and visualisation tools help to explore data?

Ans. Data exploration use statistical methods and visualisation tools to:

- Evaluate the size and quality of the data.
- Detect outliers or anomalies.
- Identify possible links between data components, files, and tables.
- Look for similarity, patterns, and outliers.
- Determine the relationships between different variables.

3. What do you mean by graphical and non-graphical data representation. Why graphical data representation is more advantageous.

Ans. **Non-Graphical Technique for Data Representation**

Non-graphical techniques include tabular and case forms. This is an older data representation format that is unsuitable for huge datasets. Non-graphical strategies are ineffective when we want to make decisions based on a set of data.

Graphical Technique for Data Representation

Graphs are commonly used to visualise statistical data using points, lines, dots, and other geometric shapes.

Advantage of graphical over non-graphical data representation: The human brain is more comfortable coping with complex and massive amounts of material when it is represented visually. Data visualisation refers to the graphical or pictorial depiction of data using graphs, charts, and other tools.

4. What is the use of autopct parameter in a pie chart?

Ans. The 'autopct' parameter displays the percentage value of each slice. Each label will show the percentage value of the corresponding slice, rounded to zero decimal place.

5. Given the following conditions, identify which chart would you use?
   i. Suitable for comparisons, particularly when there are numerous categories or negative values.
   ii. Good for displaying trends, particularly small variations or data lines that cross. They are also useful with time-series data.
   iii. Good for demonstrating the correlations and distributions of two quantitative variables. These graphs may reveal positive, negative, or no associations.
   iv. Good for demonstrating how different numbers relate to one another, such as comparing the sizes or percentages of various categories in a data set.

Ans. i. Bar charts ii. Line Graph
iii. Scatterplot iv. Pie chart

B. Long answer type questions.

1. What are the advantages of data cleaning?

Ans. Maintaining clean data will ultimately boost overall productivity and ensure the highest quality information for decision-making. The benefits include:
   i. Eliminating errors when dealing with multiple data sources.
   ii. Reducing errors, which leads to happier and satisfied clients and less-frustrated employees.
   iii. Monitoring and reporting errors to identify their sources, making it easier to correct inaccurate or corrupt data for future use.
   iv. Utilising data cleaning tools to streamline business practices and expedite decision-making.

2. What are the steps taken to clean/prepare data in the data preprocessing pipeline.

Ans. The following steps are taken to clean/prepare the data:
   - **Missing Data:** Missing data refers to the absence of certain values in the dataset, which can result from various causes. To handle missing data, strategies include removing rows or columns with missing values, imputing missing values with estimates, or utilising algorithms that can manage missing data.
   - **Outliers:** Outliers are data points that deviate significantly from the majority of the dataset, typically due to errors or uncommon occurrences. Managing outliers includes detecting and excluding them, transforming the data, or applying robust statistical techniques to minimise their influence.
   - **Inconsistent Data:** Inconsistent data, such as typographical errors or variations in data types, is rectified to ensure uniformity and coherence across the dataset.
   - **Duplicate Data:** Duplicate data is identified and eliminated to maintain data integrity and accuracy.

3. Define the following:
   - pyplot
   - Semi structured data
   - Feature Selection

Ans
- **pyplot:** Data visualisation in Python can be accomplished using the Matplotlib library. The 'pyplot' submodule of Matplotlib offers a MATLAB-like interface and includes numerous convenience functions that simplify the process of creating basic plots.
- **Semi structured data:** Text files with an apparent pattern enabling analysis e.g. HTML files.
- **Feature Selection:** This step is part of the data preprocessing pipeline. It involves choosing a subset of important features from the dataset. Feature selection is commonly done to eliminate irrelevant or redundant features from the dataset.

4. When is a scatterplot suitable?

Ans.
- It is used to track the connections between two numerical variables. When the data are seen as a whole, the dots on the plot indicate both the variable's value and any trends.
- The scatterplot is a valuable tool for calculating correlation. Variable relationships can be categorised in a variety of ways, including positive or negative, strong or weak, linear or nonlinear.
- This graph's strength lies in its ability to clearly depict trends, clusters, and relationships within datasets.

5. What do understand by the following terms:
   - Data Literacy
   - Data Representation
   - Statistics

Ans.
- **Data Literacy:** Data literacy involves the ability to find and use data proficiently. This encompasses skills like collecting data, organizing it, checking its quality, analysing it, understanding the results, and using it ethically.
- **Data Representation:** It is defined as a technique for presenting enormous amounts of data in a way that allows the user to quickly and easily interpret the most relevant information. This can be done using graphical and non-graphical techniques.
- **Statistics:** Statistics is "the discipline concerned with the collection, organisation, analysis, interpretation, and presentation of data." To accomplish this goal, statisticians summarise a significant amount of data in a compact format that yields relevant results.

C. Competency-based/Application-based questions:

Coding & Computational Thinking

In a training session the employees were given the following table:

Student	Hours Studied	Exam Score
A	2	65
B	3	70
C	4	75
D	5	80
E	6	85
F	7	90
G	8	95

The employees were asked to create a visualisation for a presentation to the Principal showing the effect of hours put to study and the corresponding exam score of the students. Which visualisation tool would you use to display the data effectively.

Ans. The data is well suited for a scatter plot. So a scatter plot will effectively show how the data relates to the final exam score recorded for various students.

Code to present the given data in the form of scatterplot in Python is:

```
import matplotlib.pyplot as plt
# Data from the table
students = ['A', 'B', 'C', 'D', 'E', 'F', 'G']
hours_studied = [2, 3, 4, 5, 6, 7, 8]
exam_scores = [65, 70, 75, 80, 85, 90, 95]

# Creating a scatter plot
plt.scatter(hours_studied, exam_scores, color='blue', marker='o')

# Adding titles and labels
plt.title('Hours Studied vs. Exam Scores')
plt.xlabel('Hours Studied')
plt.ylabel('Exam Score')

plt.show()
```

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. What does AI involve transforming raw data into?
   a. Useless information ○ b. Actionable information ○
   c. Simple statistics ○ d. Complex equations ○

2. What skills does data literacy include?
   a. Ignoring data ○ b. Proficiently finding and using data ○
   c. Only collecting data ○ d. Discarding data ○

3. What determines the quantity of data required in a machine learning model?
   a. The programming language used ○ b. The number of features in the dataset ○
   c. The design of the website ○ d. The user's preferences ○

4. Which data collection method is used to determine cause-and-effect relationships by manipulating factors to see how they affect outcomes?
   a. Interview ○ b. Survey ○
   c,. Experiment ○ d. Observation ○

5. In which level of measurement can the distance between attributes be measured, but there is no true zero point?
   a. Nominal ○ b. Ordinal ○
   c. Interval ○ d. Ratio ○

6. Which level of measurement involves ordering categories, but the intervals between them are not equal?
   a. Nominal ○ b. Ordinal ○
   c. Interval ○ d. Ratio ○

7. Which characteristic is shared by both interval and ratio scales, but not by nominal or ordinal scales?
   a. Order/Ranking ◯ b. Equal Intervals ◯
   c. Meaningful Zero Point ◯ d. Categorical Data ◯

8. Why is the median considered a better measure of central tendency than the mean in a skewed distribution?
   a. It is easier to calculate. ◯ b. It is not affected by extreme values. ◯
   c. It takes into account all values in the dataset. ◯ d. It is always a whole number. ◯

9. In a dataset with multiple modes, maximum how many modes can be there?
   a. One ◯ b. Two ◯
   c. Three or more ◯ d. Zero ◯

10. What does a line graph represent?
   a. Categorical data ◯ b. Discrete data ◯
   c. Continuous data ◯ d. Nominal data ◯

11. What type of data is best represented using a bar chart?
   a. Continuous data ◯ b. Categorical data ◯
   c. Ordinal data ◯ d. Nominal data ◯

12. Which metric is commonly used for evaluating classification problems?
   a. Mean Squared Error (MSE) ◯ b. Accuracy ◯
   c. R-squared ◯ d. Root Mean Squared Error (RMSE) ◯

13. Which technique divides the dataset into two sets: Training and Testing?
   a. Train-Test Split ◯ b. Cross-Validation ◯
   c. Error Analysis ◯ d. Feature Selection ◯

14. What is the term used for data points that deviate significantly from most of the dataset i.e. extreme values?
   a. Normals ◯ b. Anomalies ◯
   c. Variations ◯ d. Outliers ◯

B. Fill in the blanks.

1. The main goal of AI data analysis is to extract valuable information that can aid in _____ and drawing conclusions.
2. _____ is a specialized tool used in surveys – is a collection of questions meant to gather information from respondents.
3. _____ use statistical methods and visualisation tools to look for similarity, patterns, and outliers.
4. _____ divides the dataset into two equal halves, with half of the values being less than the median and half being greater.
5. A _____ is a strong tool for representing continuous data on a numbered axis.
6. _____ converts continuous data into discrete categories.
7. The size of the _____ data is smaller than the _____ data because it used to evaluate the performance of the model.
8. _____ ensures consistent model performance across different data subsets.

C. State whether the following statement is True or False.

1. The method used to measure a collection of data is known as the level of measurement. _____
2. Secondary data is considered valuable because it is tailored to the specific research or business needs. _____

3. Nominal variables are like labels or categories like car brands or seasons. _____

4. Interval level data shares similarities with ordinal data as it maintains a clear order, but it differs in that the differences between values can be measured. _____

5. Mode represents the sum of all values in a dataset divided by the number of values. _____

6. Data points with a low standard deviation are near to the mean. _____

7. Matplotlib is non-customisable, giving users little control over the appearance of their plots. _____

8. Scatterplot is referred to as horizontal bar chart. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. Define Test Data.
2. Explain the levels of ordinal data with example.
3. What is a pie chart? Explain with example.
4. What are the advantages of using matplotlib for data visualisation?
5. In which conditions should you choose to use median in your calculations?

B. Long answer type questions.

1. Discuss different sources of secondary data.
2. What is structured, semi-structured and unstructured data? Give examples.
3. Explain the Data Transformation process.
4. List any 3 functions of Matplotlib. Also mention which graph they help to plot.
5. Explain the process of Data Preprocessing.

C. Competency-based/Application-based questions:

Coding & Computational Thinking

1. Computer science teacher in Gyan Vidya School invited students to try data visualisation for the following data. Can you suggest a suitable method for visualising this data. Also give the Python code required for the same.

Month	Max. Temperature (°C)
Jan	19.50
Feb	23.00
Mar	27.95
Apr	34.97
May	37.66
Jun	36.26
Jul	32.90
Aug	31.77
Sep	31.73
Oct	30.92
Nov	27.33
Dec	22.77
Average	29.73

2. Your teacher has asked you to compare the test scores of two classes to understand which class has more consistent performance. Class A's test scores are [85, 90, 92, 88, 91], and Class B's test scores are [78, 82, 85, 79, 81]. Calculate the variance and standard deviation for both classes to determine which class has more consistent scores.

Ai In Life

Interdisciplinary

Statistics is the heart of data science, helping to evaluate, transform, and predict data. So, if you want to excel in this amazing field, you must first become acquainted with the relevant Statistics topics for data science. Few are mean, median, mode, correlation and standard deviation. Find out 5 more Statistics concepts you should know for a career in Data Science.

Ai Deep Thinking

Creativity and Innovativeness

Misleading interpretations resulting from inaccurate or "unclean" data can influence flawed business strategies and decision-making processes. Such misinterpretations may lead to embarrassing situations during reporting meetings when the inadequacy of the data becomes apparent. It is crucial to establish a culture of high-quality data within your organisation to avoid such scenarios. To achieve this, it is necessary to document the tools employed to foster this culture and define the standards and criteria for data quality that are meaningful to your organisation. Find out any five characteristics of quality data.

Ai Lab

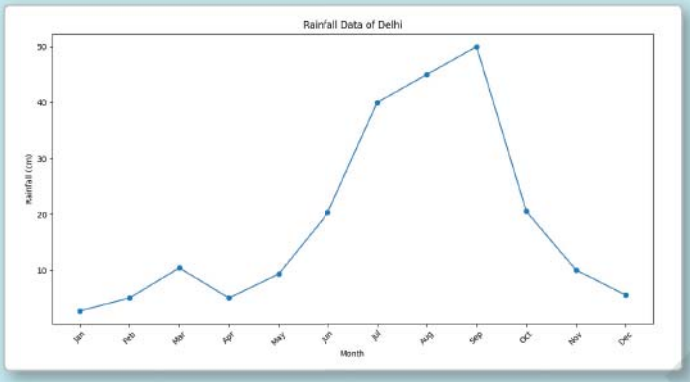
Coding & Computational Thinking

1. **Construct a simple line graph to represent the rainfall data of Delhi as shown in the table below using Python:**

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rainfall (cm)	2.7	5	10.4	5	9.3	20.3	40	45	50	20.5	10	5.5

Ans.

```
import matplotlib.pyplot as plt
# Data
months = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep',
'Oct', 'Nov', 'Dec']
rainfall = [2.7, 5, 10.4, 5, 9.3, 20.3, 40, 45, 50, 20.5, 10, 5.5]
# Create line graph
plt.plot(months, rainfall, marker='o', linestyle='-')
# Add title and labels
plt.title('Rainfall Data of Delhi')
plt.xlabel('Month')
plt.ylabel('Rainfall (cm)')
# Rotate x-axis labels for better readability
plt.xticks(rotation=45)
# Show plot
plt.show()
```



2. **You ask your friends which 'food item' they like best. Here is what you found:**

Pizza	Burger	Momos	Rolls	French Fries
3	5	6	2	4

Create a pie chart of the above data and answer the following questions after studying the graph.

i) Which food item is liked the most?

ii) Which food item is liked the least?

iii) Which food items together constitute 50% of the graph?

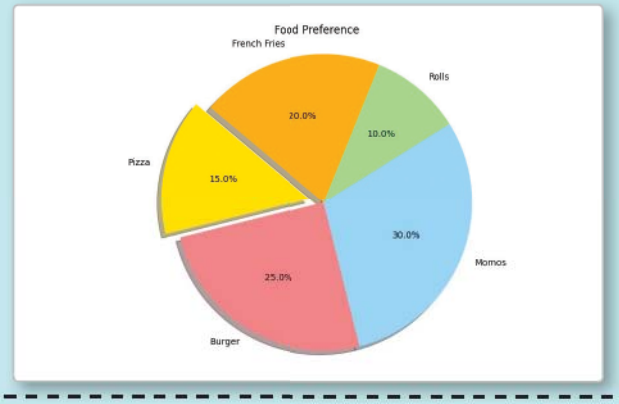
Ans.

```
import matplotlib.pyplot as plt
# Data
labels = ['Pizza', 'Burger', 'Momos', 'Rolls', 'French Fries']
sizes = [3, 5, 6, 2, 4]
colors = ['gold', 'lightcoral', 'lightskyblue', 'lightgreen', 'orange']
explode = (0.1, 0, 0, 0, 0)  # explode the 1st slice (Pizza)
# Create pie chart
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=140)
# Add title
plt.title('Food Preference')
# Equal aspect ratio ensures that pie is drawn as a circle.
plt.axis('equal')
# Show plot
plt.show()
```

i) Momos

ii) Rolls

iii) Momos & French Fries

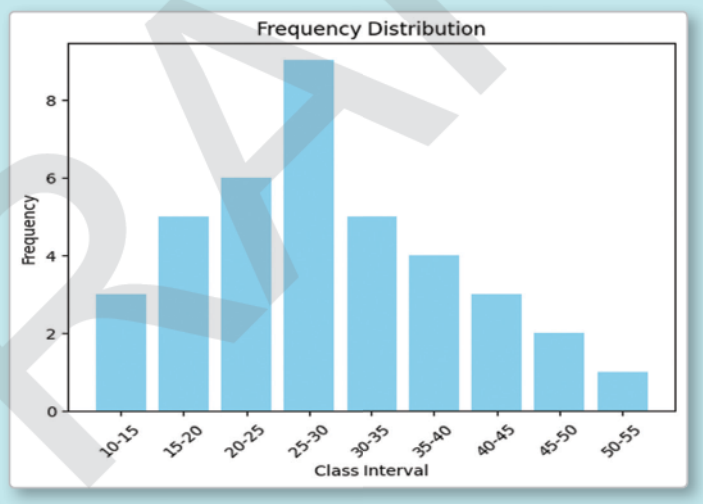


3. **The following data is stored in a csv file named 'data.csv'. Prepare a histogram of the following distribution using Python:**

Class interval	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55
Frequency	3	5	6	9	5	4	3	2	1

Ans.

```
import matplotlib.pyplot as plt
import pandas as pd
# Load data from CSV file
data = pd.read_csv('data.csv')
# Extract class intervals and frequency from the loaded data
class_intervals = data['Class interval']
frequency = data['Frequency']
# Create histogram
plt.bar(class_intervals, frequency, color='skyblue')
# Add title and labels
plt.title('Frequency Distribution')
plt.xlabel('Class Interval')
plt.ylabel('Frequency')
# Rotate x-axis labels for better readability
plt.xticks(rotation=45) #The x-axis labels are rotated by 45 degrees for bet-
ter readability.
# Show plot
plt.show()
```



4. **My Netflix viewing capacity is given in a table as follows:**

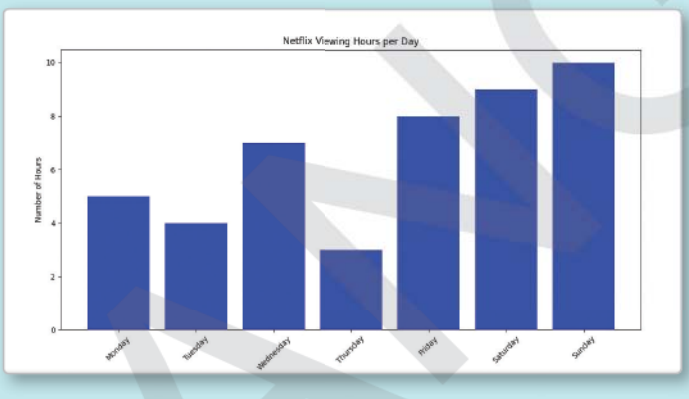
Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
No. of hours	5	4	7	3	8	9	10

Prepare a bar chart using Python.

Ans.

```
import matplotlib.pyplot as plt


# Data
```

```
days = ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday',
'Sunday']
hours = [5, 4, 7, 3, 8, 9, 10]

# Create bar chart
plt.bar(days, hours, color='blue')

# Add title and labels
plt.title('Netflix Viewing Hours per Day')
plt.xlabel('Day')
plt.ylabel('Number of Hours')
plt.xticks(rotation=45)

# Show plot
plt.show()
```



Refer the above codes through this link or scan the QR code
https://colab.research.google.com/drive/1Pp4u5Oh4y_J0YkZTG_GlUMrgOVHRkruG?usp=sharing

Answers

Exercise (Section A)

A. 1. b 2. d 3. c 4. b 5. b 6. d 7. d 8. a 9. b 10. d
11. c 12. b 13. c 14. d 15. b

B. 1. Matplotlib 2. primary, secondary 3. Satellite Data Tracking 4. Statistical analysis
5. one another 6. larger 7. DataFrame
8. Normalisation

C. 1. True 2. True 3. False 4. False 5. True
6. True 7. True 8. False

Read the blog-Data Preprocessing in Data Mining: A Hands On Guide -
https://www.analyticsvidhya.com/blog/2021/08/data-preprocessing-in-data-mining-a-hands-on-guide/

Answer the following questions.

1. What is meant by noisy data?

2. Mention any 2 techniques listed in the blog to handle missing values.

3. List 2 problems to be considered during data integration.

4. What is Normalisation?

UNIT-6

MACHINE LEARNING ALGORITHMS

Learning Outcomes

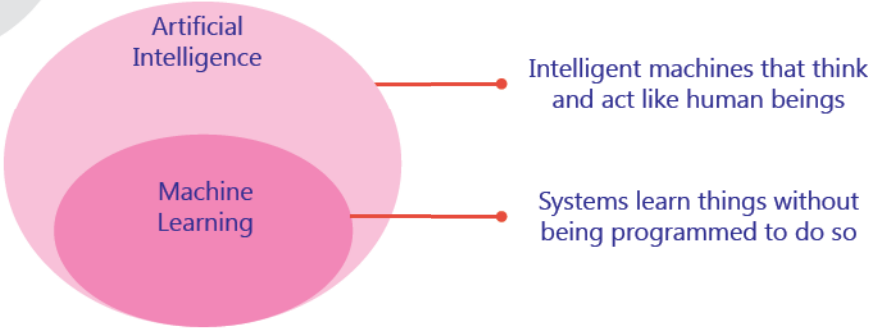
- Machine Learning in a Nutshell
- Regression
- Classification
- K-Means Clustering
- Types of Machine Learning
- Correlation
- Unsupervised Learning—Clustering
- Why is Clustering Unsupervised?

Machine learning (ML) is a subset of Artificial Intelligence that allows computers to learn and make judgments based on data. It involves algorithms that can identify patterns, anticipate outcomes, and improve over time without being specifically coded for specific tasks. Here's a real-world example, assume you have a personal assistant who can foresee your wants, provide solutions, and even forecast issues before they arise. This is what machine learning can achieve, making it a valuable tool in today's technologically advanced world.

Machine learning has the capacity to extract information from huge datasets and perform complicated operations that transform various industries ranging from healthcare and banking to transportation and entertainment. As researchers continue to invent and improve these algorithms, machine learning's impact on society is set to grow, ushering in a period of extraordinary creativity and opportunity.

Machine Learning in a Nutshell

Machine learning is a subdomain of Artificial Intelligence (AI) that focuses on developing systems that learn or improve performance based on the data they ingest. It is an application of Artificial Intelligence (AI) that enables systems to learn and improve automatically from experience without the need for explicit programming. It focuses on the development of computer programs that can access data and use it to learn for itself. Data is critical for machine learning to work. The more data the machine is given (assuming that the data is reliable), the more accurate its prediction is. Artificial Intelligence is a broad word that refers to systems or machines that resemble human intelligence. A crucial distinction between Artificial Intelligence and Machine Learning is that all ML is AI, but not all AI is ML.



Machine learning algorithms can learn from different kinds of information, such as pictures, text, sensor readings, and past data, by figuring out patterns in the data to guess or decide things. Additionally, some common machine learning methods such as decision trees, neural networks, and support vector machines enable this learning process.

Features of Machine Learning

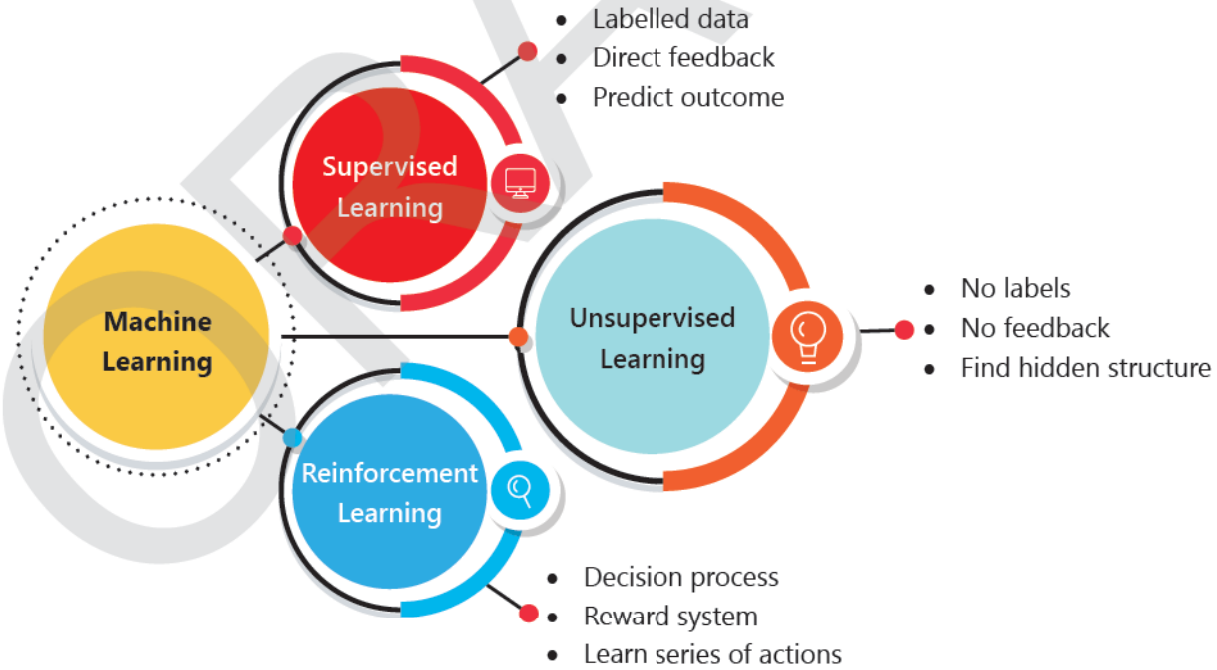
Some key features of machine learning are as follows:

- ML interprets, analyses, and processes data to address real-world problems.
- It learns from data and enhances its performance over time.
- The technology facilitates automation and prediction based on learned patterns.
- It is the prevailing approach in contemporary AI.
- It employs data analysis, training, and sometimes human review to refine its capabilities.
- Unlike traditional programming, it doesn't rely on predefined rules but learns from examples and experiences.
- It powers a wide range of applications across industries, from healthcare and finance to autonomous vehicles and recommendation systems.

However, ML is not without its challenges. Overfitting, in which models specialise too much on training data, can result in poor performance on fresh data. Bias in training data can lead to skewed predictions, and some models are difficult to understand, serving as black boxes. Despite these hurdles, machine learning (ML) converts data into knowledge, allowing computers to learn, adapt, and make autonomous judgments.

Types of Machine Learning

Machine learning can be divided into three primary categories, each distinguished by its learning approach and nature of the input data:



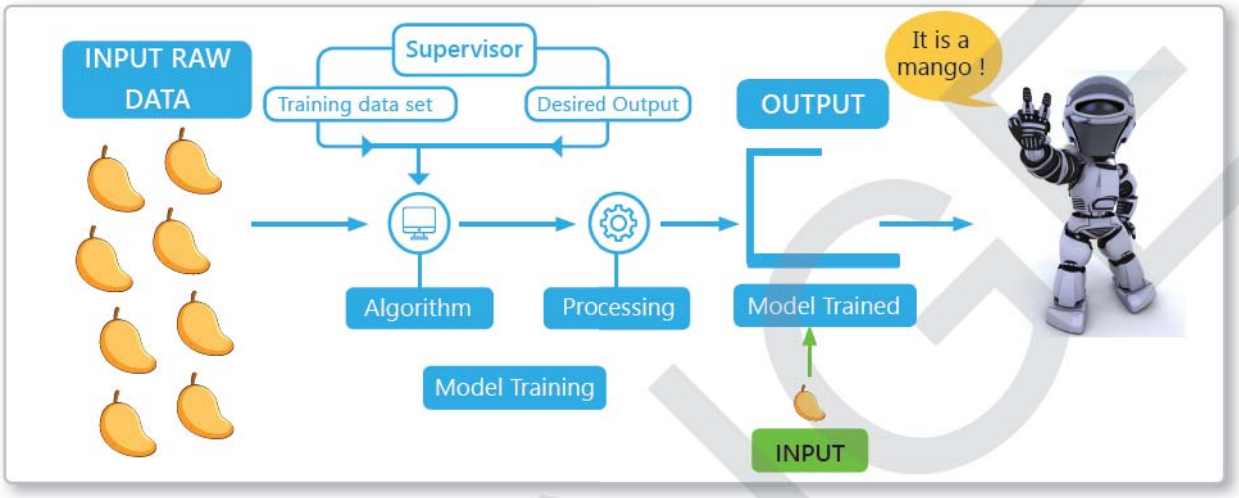
Supervised Learning

Supervised learning is a type of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.

In supervised learning, the algorithm learns from labelled data, where each training example is paired with a corresponding target label. The goal is to learn a mapping from input variables to output labels. During training, the algorithm adjusts its parameters to minimise the difference between predicted and actual labels.

Building, expanding, and successfully implementing accurate supervised machine learning models requires time and technical expertise from a team of highly trained data scientists. In the real-world, supervised learning can be used for risk assessment, image classification, fraud detection, spam filtering, etc. Take a look at the examples below:

Example 1:



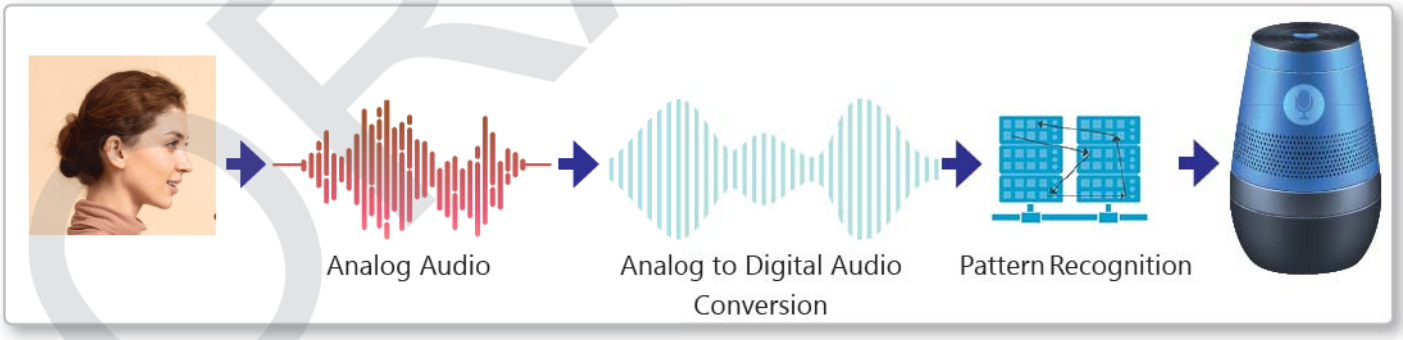
Step 1: You provide the system with images of mangoes and tag them as mangoes. This type of input is referred to as labelled data.

Step 2: The model learns from the labelled data and next time you ask it to identify a mango, it can do it easily.

That's exactly how supervised learning works.

Example 2:

Many voice assistants, including Apple's Siri and Amazon's Alexa, use supervised learning algorithms to process and interpret spoken instructions. The algorithms are trained on a dataset of labelled speech data (transcribed speech and text), which they then use to transcribe and interpret spoken commands.

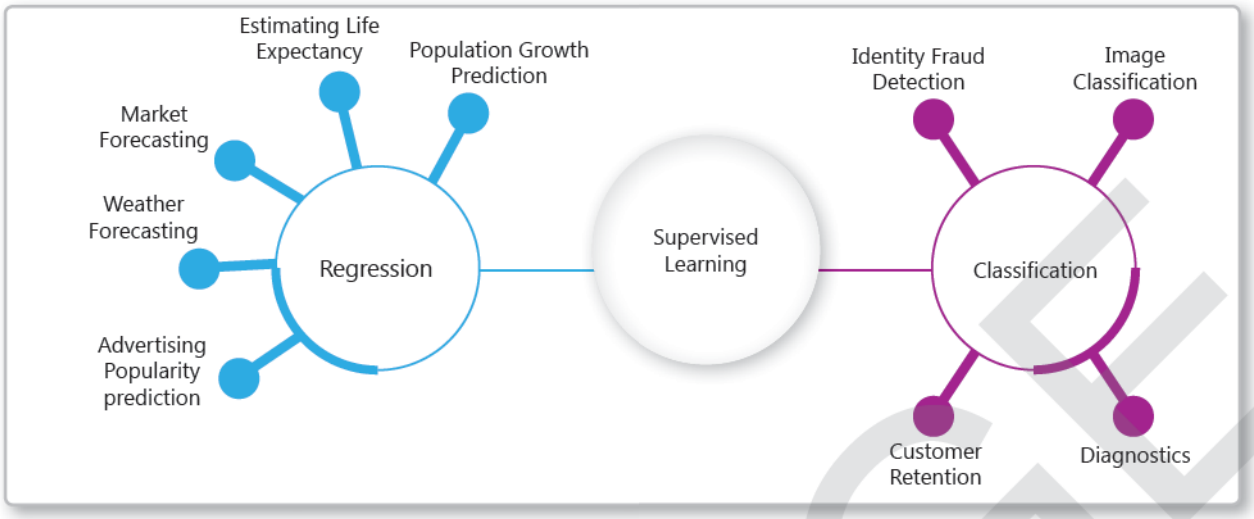


Supervised Learning Algorithms and Their Use

Supervised learning involves two primary algorithms: regression and classification.

- **Regression algorithms** create a mapping function from the input data, allowing us to predict continuous outcomes. They are used if there is a relationship between the input variables and the output variables. For instance, predicting house prices based on features like size and location is a task for regression.
- **Classification algorithms**, on the other hand, involve creating a function that assigns data points to specific categories. They are used when the output variable is categorical, implying there are two classes such as Yes-No, Male-Female, True-false, etc.

For example, classifying emails as spam or not spam uses a classification function to determine the appropriate category for each email. Following are the uses of Supervised Learning:



Advantages of supervised learning:

- With the help of supervised learning, the model can predict the output on the basis of prior experiences.
- In supervised learning, we can have an exact idea about the classes of objects.
- Supervised learning model helps us to solve various real-world problems such as **fraud detection**, **spam filtering**, etc.

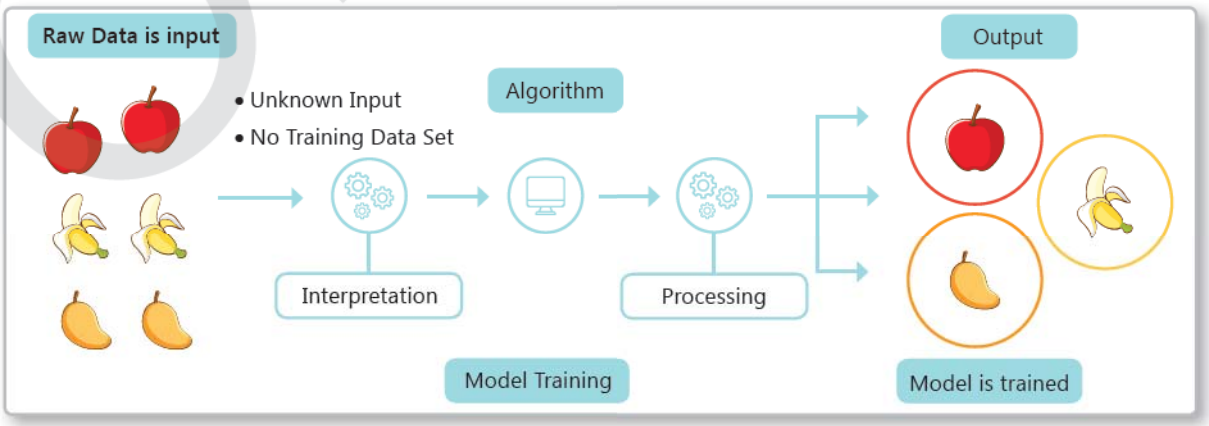
Disadvantages of supervised learning:

- Supervised learning models are not suitable for handling the complex tasks.
- Supervised learning cannot predict the correct output if the testing data is different from the training dataset.
- Training requires a lot of computational time.
- In supervised learning, we need enough knowledge about the classes of object.

Unsupervised Learning

As the name suggests, unsupervised learning is a machine learning technique in which models are not supervised using training dataset. The machine learns through observation and finds patterns in data. The system will explore the data and draw inferences from the data set to describe the hidden patterns in the unlabelled data. Unsupervised machine learning algorithms are used when the information used to train is neither classified nor labelled.

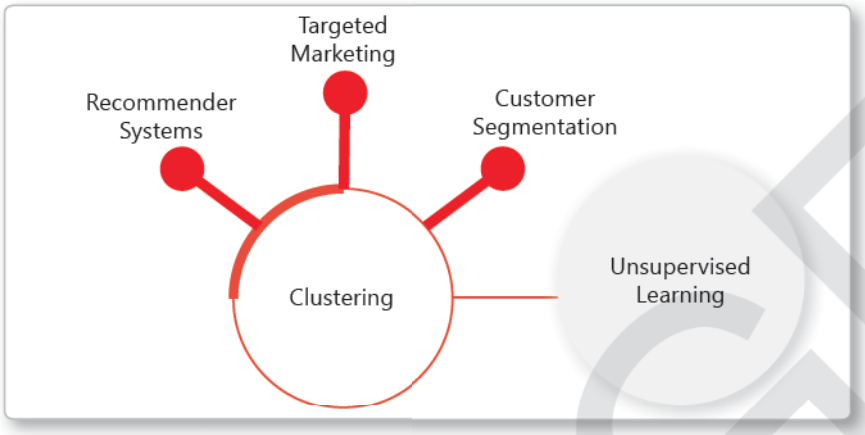
For example, if somebody gives you a basket full of different fruits and asks you to separate them, you will probably do it based on their colour, shape, and size, right? Unsupervised learning works in the same way. For example,



Unsupervised Learning Algorithm and Its Uses

Clustering and **Dimensionality Reduction** are common algorithms in unsupervised learning.

Clustering is a machine learning approach where the machine partitions the dataset into different clusters or categories based on similar characteristics. The uses of clustering algorithms are:



Dimensionality Reduction is a technique used to reduce the number of features or variables in a dataset while preserving the most important information. It is particularly useful when dealing with high-dimensional data, where the number of features is large relative to the number of samples. **Dimensionality Reduction** methods aim to simplify the dataset, making it easier to visualise, analyse, and model while also reducing computational complexity.

Advantages of Unsupervised Learning:

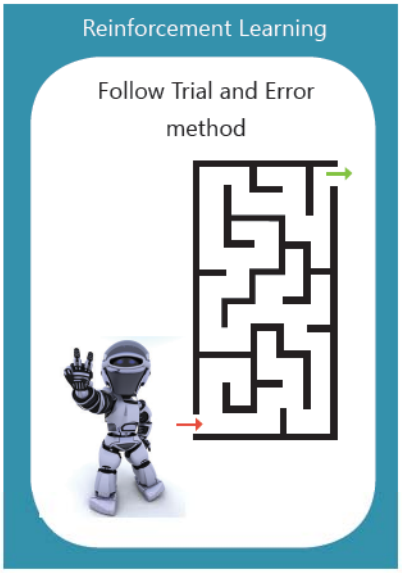
- Unsupervised learning is used for more complex tasks as compared to supervised learning because, in unsupervised learning, we don't have labelled input data.
- Unsupervised learning is preferable as it is easy to get unlabelled data in comparison to labelled data.

Disadvantages of Unsupervised Learning:

- Unsupervised learning is intrinsically more difficult than supervised learning as it does not have corresponding output.
- The result of the unsupervised learning algorithm might be less accurate as input data is not labelled, and algorithms do not know the exact output in advance.

Reinforcement Learning

Reinforcement Learning(RL) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences. The agent takes actions and observes the outcomes, receiving feedback in the form of rewards or penalties. Over time, through repeated trials and adjustments to its strategy, the agent refines its decision-making process to achieve better performance and maximise its cumulative reward. This iterative approach allows the agent to learn optimal behaviours without explicit instructions or supervision.



Examples of Reinforcement Learning: chess game, text summarisation

Regression

Regression is a Supervised Machine Learning algorithm used to analyse the relationship among dependent variable (target) and independent variable (predictor). The objective is to determine the most suitable function that characterises the connection between these variables.

It predicts the output values based on input values. It is mainly used for weather forecasting, finding the causal-effect relationship between variables and time series modelling.

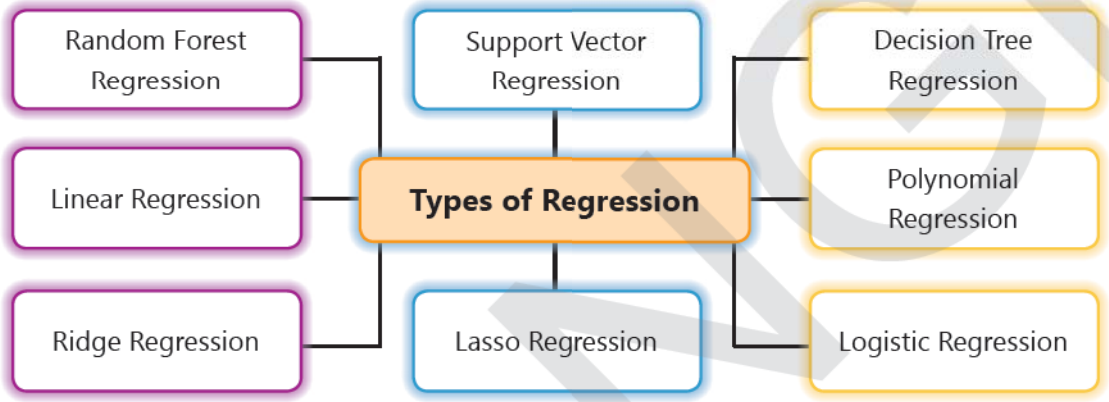
In regression tasks, there are two kinds of variables being studied: the dependent variables and the independent variables.

- **Independent variables:** Quantities that can be measured directly.
- **Dependent variables:** Quantities whose value depends on independent variables.

As the independent variable is adjusted, the level of the dependent variable will vary. The dependent variable is the variable under study, and it is the variable that the regression model tries to predict. In the linear regression task, each observation is made up of the value of the dependent variable and the value of the independent variable.

Regression is basically used when the dependent variable is of a continuous data type. The independent variables, on the other hand, can be of any data type—continuous, nominal/categorical etc.

There are several types of regression analysis, which are as follows:



Linear Regression—Finding the Line

When we make a distribution in which there is an involvement of more than one variable, then such an analysis is called Regression analysis. Regression generally focuses on predicting the value of the variable that is dependent on the other variable. Let us consider two variables *x* and *y*.

y – Regression or Dependent Variable

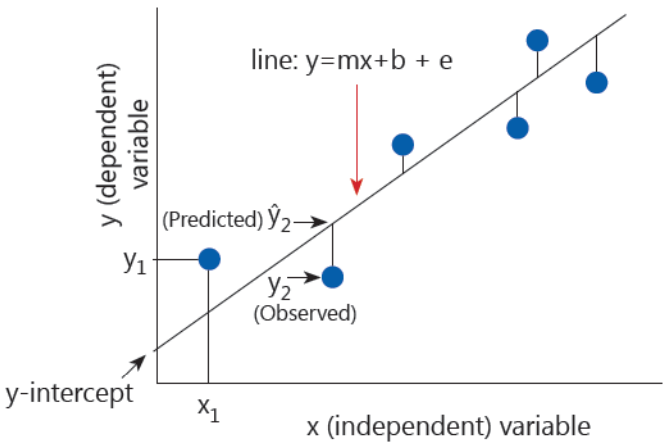
x – Independent Variable or Predictor

Therefore, if we use a simple linear regression model where y depends on x, then the regression line of y on x is:

$y = mx + b + e$

where,

- **x** is the independent variable.
- **y** is the dependent variable.
- **m** is the slope of the line.
- **b** is the y-intercept.
- **e** is the residual error and represents y(observed) - ŷ(predicted) or $(y_2 - \hat{y}_2)$

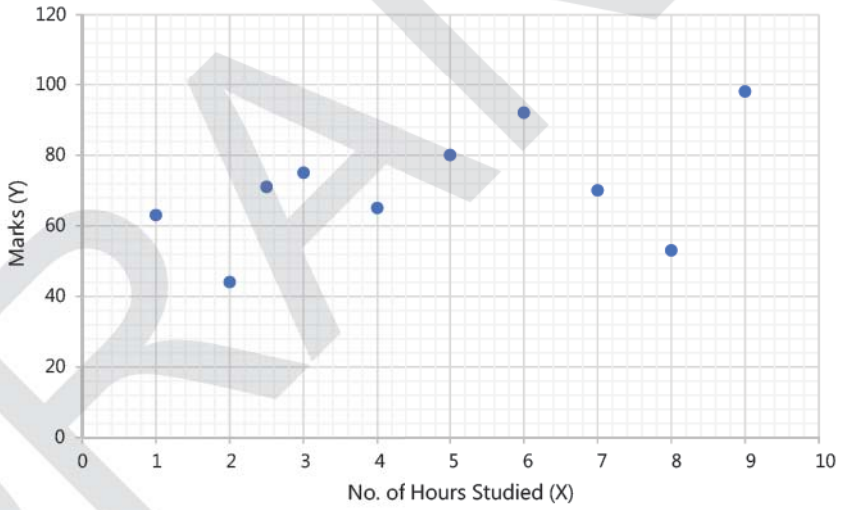


Least Squares Method—Finding the Line of Best Fit

Consider the following example, where marks of 10 students are shown, which they scored after a certain number of hours of study:

No. of Hours Studied	Marks
2	44
9	98
5	80
3	75
7	70
1	63
8	53
6	92
2.5	71
4	65

Assuming **No. of Hours Studied** as x and **Marks** as y, let us learn to plot the above data on a Scatterplot using MS Excel.



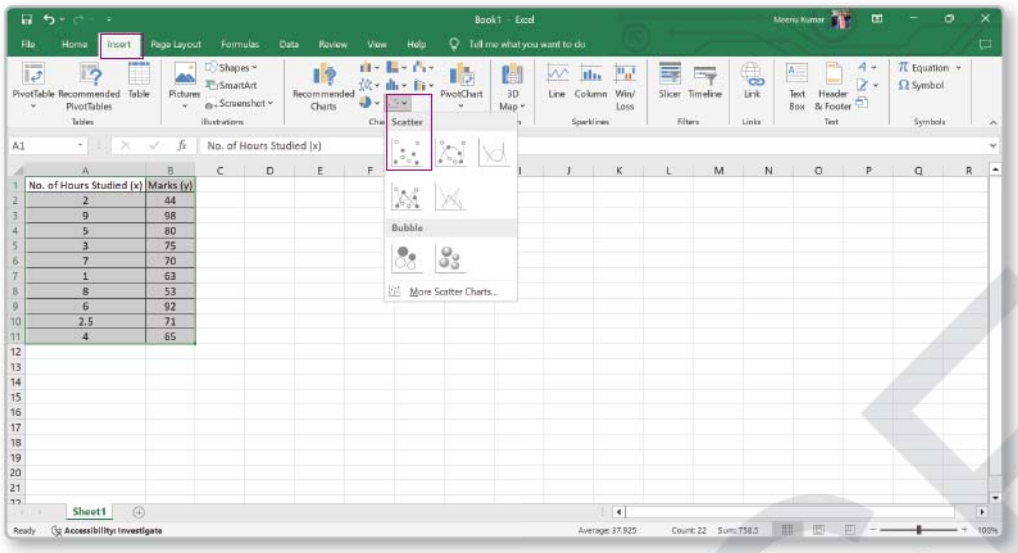
We will also try to find the line that best fits the data i.e. the line that passes close to most of the data points. This line is called the '**Line of Best Fit**' or '**Regression Line**'.

Let us find the m (slope) and b (y-intercept) that suits that data

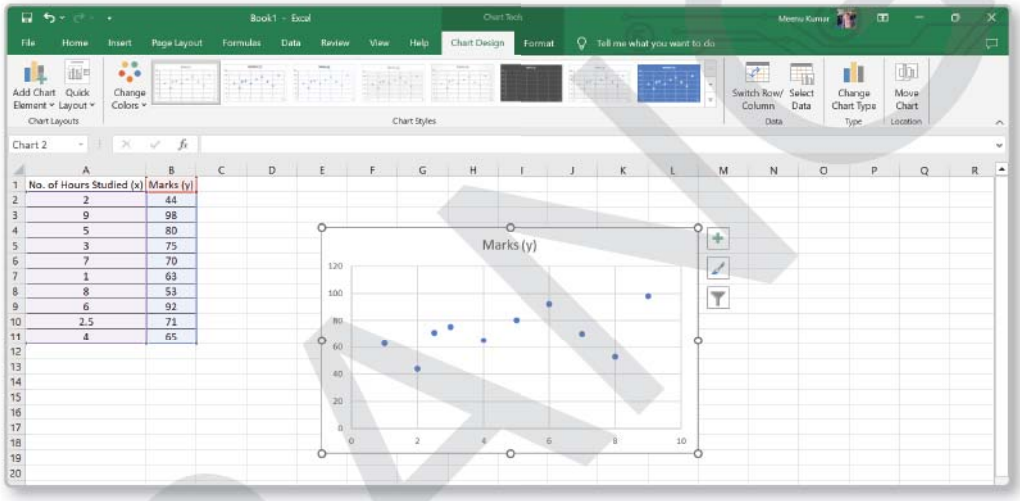
$y = mx + b + e$

Step 1: Type the given data in **Excel**.

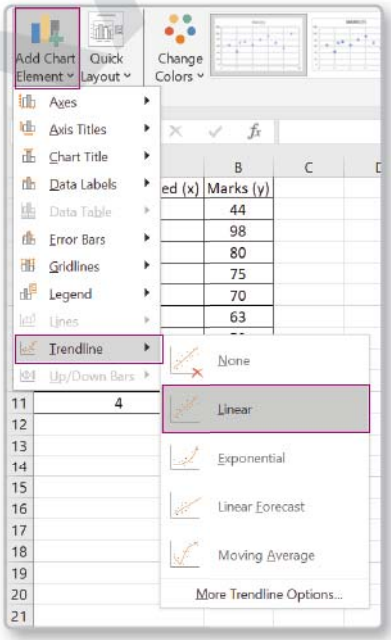
Step 2: Select x and y. Click on **Insert Chart** → select the first scatter plot option.



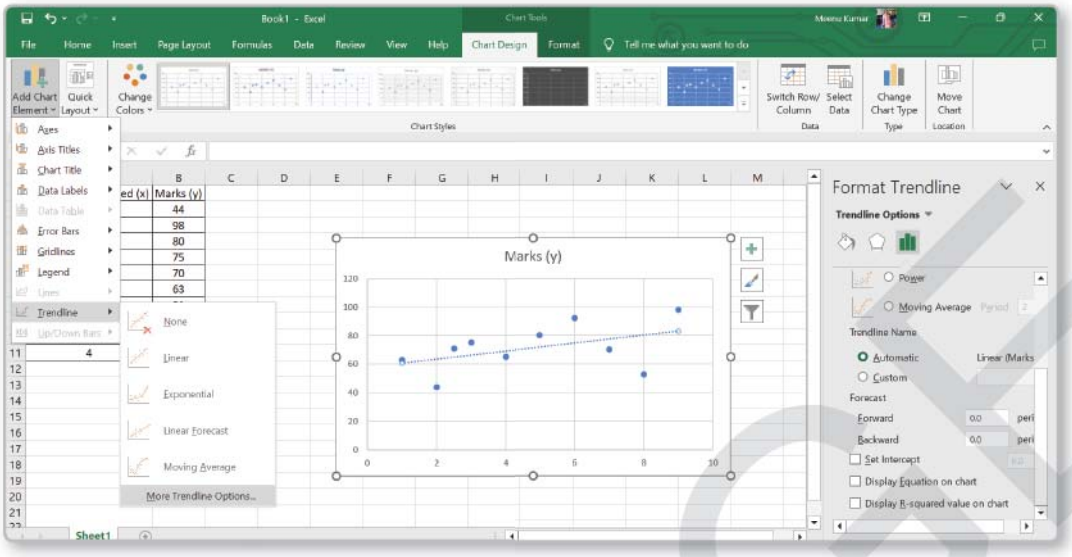
You will get a scatter plot as follows:



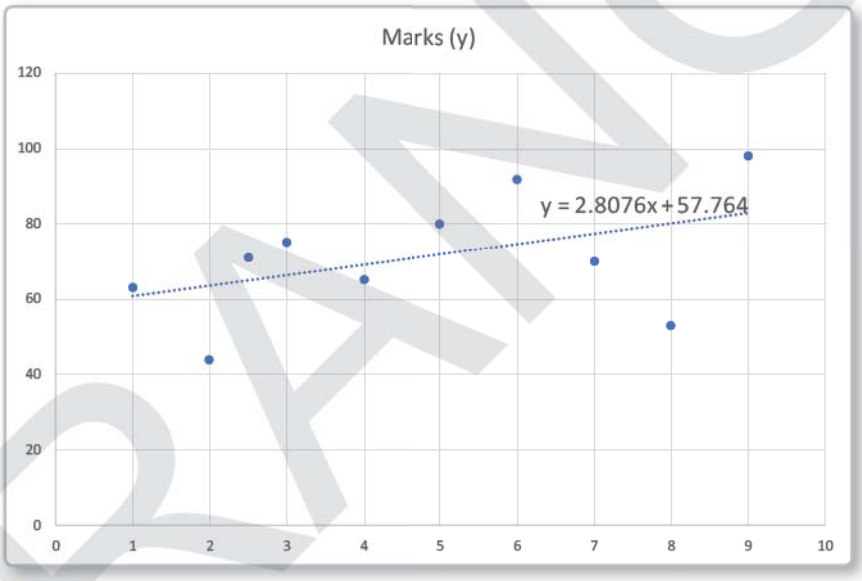
To add the Regression Line, click on **Add Chart Element → Trendline → Linear**



Next, to get the Regression Equation, click on **Add Chart Element → Trendline → More Trendline Options**. A Format Trendline dialogbox will open on the right. Click on **Display Equation on Chart** checkbox..



You will get the Regression Equation on the chart as follows:

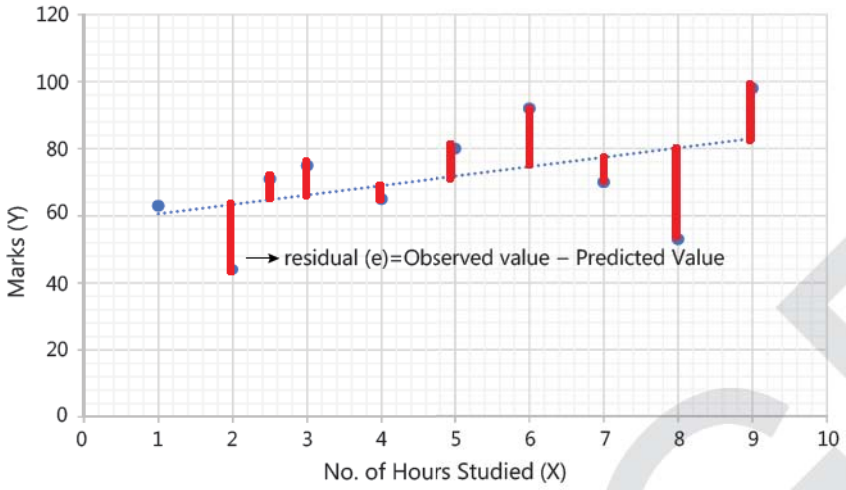


Step 3: You can also calculate the value of slope by using the slope function in Excel. Click on any cell, and type =slope((select all y values) , (select all x values). Press enter key to see the answer.

Step 4: You can also calculate the value of intercept by using the intercept function in Excel. Click on any cell, and type =intercept ((select all y values) , (select all x values). Press enter key to see the answer.

Using the regression equation, we can calculate predicted values of y using any values of x.

The vertical distance between the observed responses in the dataset and the line of best fit is called the residual error (e) as shown in the graph below:



Regression—How good is the Line?

1. Linear regression aims to find the best-fitting straight line through the points.
2. If data points are closer to the line of best fit (less residual error), it means the correlation between the two variables is higher. That means, the relationship between the two variables is strong.
3. The regression line is also called '**Line of Least Squares of Errors**' because the lower the residual errors, the better.
4. Each data point has one residual.

Digital Literacy

Video Session

Scan the QR code or visit the following link to watch the video: Linear Regression Algorithm

https://www.youtube.com/watch?v=E5RjzSK0fvY&t=527s

Brainy Fact

The least squares regression method was first published by mathematicians Legendre in 1805 and Carl Friedrich Gauss in 1809. Both used linear regression to predict the movement of planets around the sun. Gauss later published an improved method in 1821.

When Regression Analysis is Not Suitable

It's important to understand that regression analysis may not always be suitable in certain scenarios:

- **No Correlation:** If there is no correlation between the variables, meaning they change independently of each other, regression analysis will not yield meaningful insights or predictions.

- **Non-linear Relationships:** Regression is effective for modelling linear relationships but may not accurately capture more complex, non-linear relationships. In such cases, techniques like polynomial regression or non-linear regression might be more appropriate.
- **Outliers:** Outliers, or extreme data points, can disproportionately affect the regression model and lead to inaccurate predictions. It's crucial to assess the impact of outliers and consider alternative modelling approaches if necessary.
- **Violation of Assumptions:** Regression analysis depends on certain assumptions, such as linearity of relationships and absence of multicollinearity (high correlation between predictor variables). If these assumptions are violated, the regression analysis results may be unreliable.

Linear regression is a supervised learning algorithm. It makes use of one independent variable, x, to predict the outcome of a second dependent variable, y. This method finds the most accurate straight line that best describes the relationship between the dependent and the independent variables, with minimum error.

Applications of Linear Regression

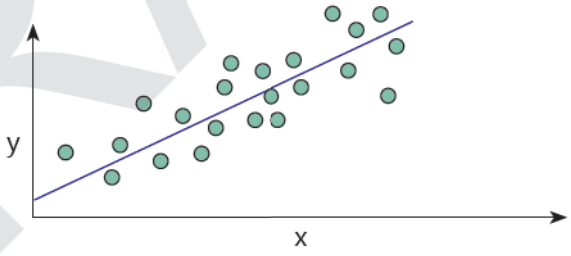
Linear regression is used in various Artificial Intelligence applications. It has its limitations, but its simplicity, interpretability, and efficiency often exceed these limitations. Real life applications of linear regression include:

- Prediction of product demand
- Sales forecasting
- Analysing the effect of price change of a service
- Predict the effect of fertiliser on crop yield
- Prediction of revenue through advertisements
- Predicting salary of a person based on number of years of experience

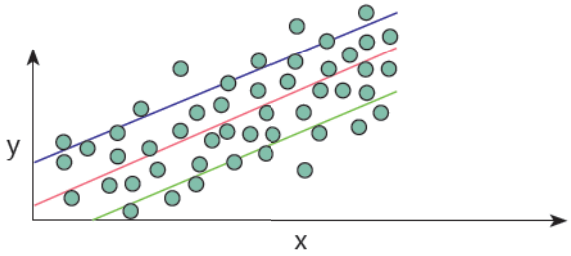
Types of Linear Regression

There are two types of Linear Regression, which are as follows:

- **Simple linear regression:** It refers to the utilisation of a single independent variable for forecasting an outcome of a numerical dependent variable.



- **Multiple linear regression:** It demonstrates a connection between two or more independent variables and the associated variables that are dependent. The variables that are independent can be continuous or categorical. This kind of regression type allows you to forecast patterns, predict potential outcomes, and forecast the effects of adjustments.



Advantage of Linear Regression

Some advantages of linear regression are as follows:

- Linear regression is a simple technique and easy to implement.
- Efficient to train the machine on this model.

Disadvantages of Linear Regression

Some disadvantages of linear regression are as follows:

- Regression analysis is sensitive to outliers as these can have a great impact on the analysis.
- It is quite prone to overfitting. (Overfitting means that the training of the model on data is just too good and the test sample size is quite small).

Ai Reboot

1. State the two types of Regression.
2. How many variables are used in linear regression?
3. State the equation of the line of best fit.
4. Why is it called the line of best fit?
5. State two applications of regression.

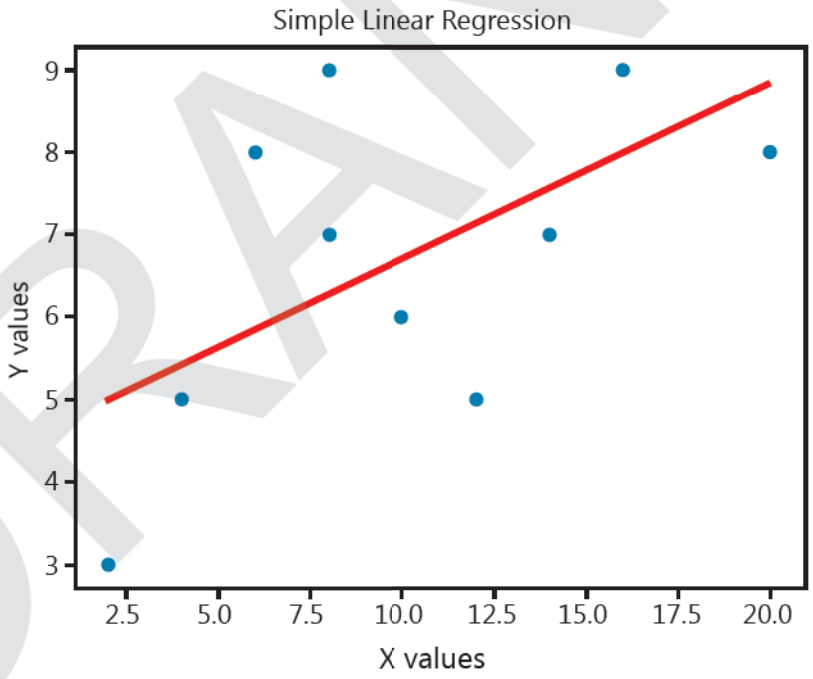
For Advanced Learners

Program 1: To demonstrate the use of simple linear regression in Python

```
import numpy as np
import matplotlib.pyplot as plt

# Updated sample data with more values
data_x = np.array([2, 4, 8, 6, 8, 10, 14, 12, 16, 20])
data_y = np.array([3, 5, 7, 8, 9, 6, 7, 5, 9, 8])

# Calculate mean and standard deviation
mean_x = np.mean(data_x)
mean_y = np.mean(data_y)
std_x = np.std(data_x)
std_y = np.std(data_y)

# Calculate covariance and slope
cov = np.sum((data_x - mean_x) * (data_y - mean_y)) / (len(data_x) - 1)
slope = cov / (std_x**2)

# Calculate y-intercept (b)
intercept = mean_y - slope * mean_x
```

```
# Predicted values
y_pred = slope * data_x + intercept

# Plot data and regression line
plt.scatter(data_x, data_y)
plt.plot(data_x, y_pred, color='red')

# Add labels and title
plt.xlabel('X values')
plt.ylabel('Y values')
plt.title('Simple Linear Regression')

# Show the plot
plt.show()

# Print slope and intercept
print("The slope of the regression line is: {:.2f}".format(slope))
print("The intercept of the regression line is: {:.2f}".format(intercept))
```

Output:



The slope of the regression line is: 0.21

The intercept of the regression line is: 4.56

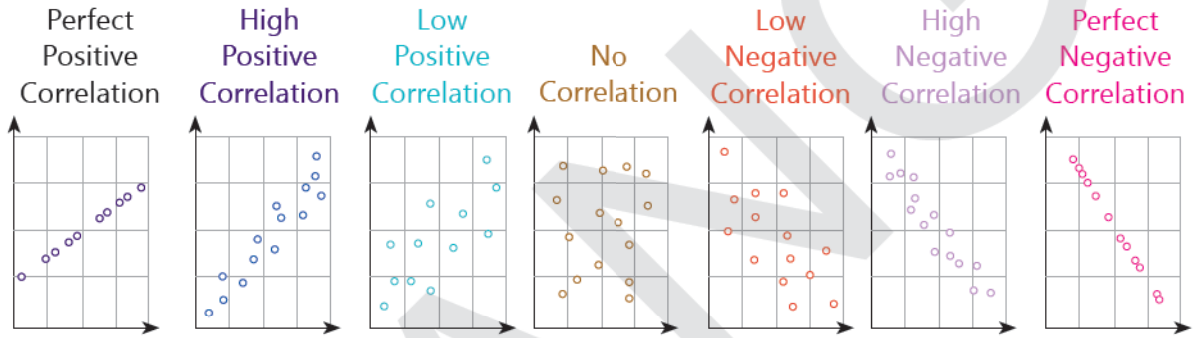
In Program 1,

- The program imports NumPy for numerical calculations and Matplotlib's pyplot module for plotting.
- Sample data for both the independent variable (x) and the dependent variable (y) is provided.
- Mean, standard deviation, covariance, and slope of the data are computed to understand their relationships.

- Using the slope and mean values, the y-intercept of the regression line is calculated.
- Predicted y-values are calculated based on the linear equation formed by the slope and y-intercept.
- The data points and the regression line are plotted on the same graph to visualise the relationship between x and y.
- Labels for x and y axis are added to the plot as well as, a title is provided to describe the purpose of the plot. The plot is displayed for visualisation.
- The estimated slope and intercept values are printed to provide insights into the relationship between the variables.

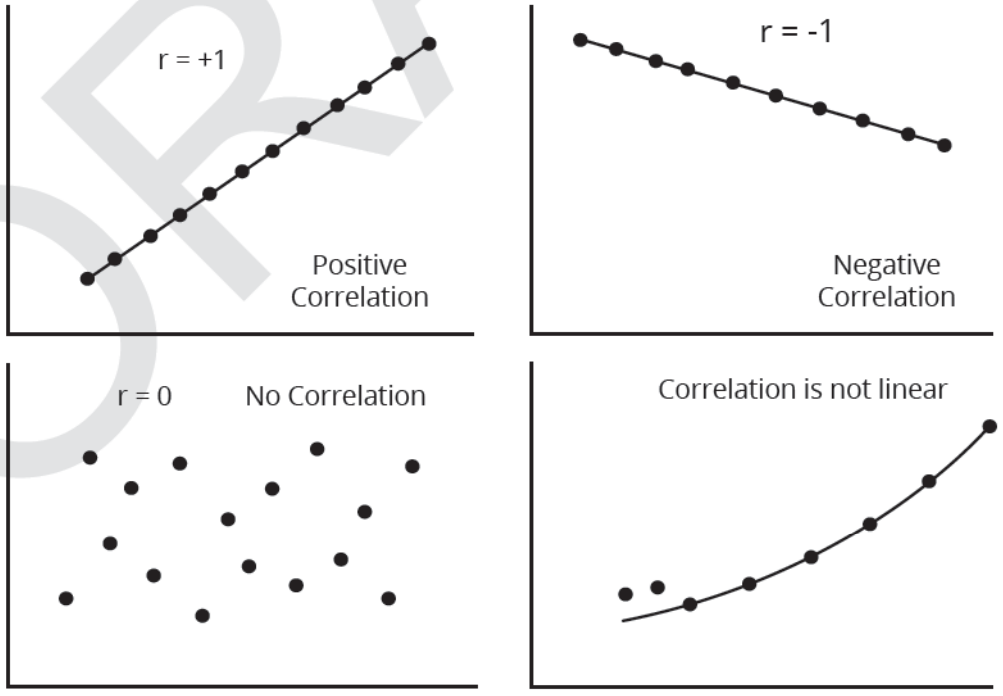
Correlation

The word **correlation** is used in daily life to denote some forms of association. We might say that we have noticed a correlation between smog and asthma attacks. However, in statistical terms, we use correlation to express an association between two quantitative variables. It measures the strength or degree of relationship between two variables. The relationship may be causal. We also presume that the association is linear, i.e., one variable increases or decreases a set amount for a unit increase or decrease in the other.



Types of Correlation

There are four types of correlations:



- **Positive correlation:** Positive correlation is the relationship between two variables, in which both variables have a linear relationship. As one variable increases/decreases, the second variable too increases/decreases. For example, when fuel prices increase, prices of airline tickets also increase.
- **Negative correlation:** Negative correlation is the relationship between two variables, where one variable increases as the second variable decreases, and vice versa. For example, more exercising leads to a decrease in body weight.
- **No correlation:** No correlation means that there is no relationship between two variables. If the value of a variable is changed, another variable is not affected. For example, shirt size and monthly expense, body weight and intelligence, etc.
- **Non-linear correlation:** A non-linear correlation is a correlation in which the relationship between variables may not always be a straight line and all the points of a scatter plot are tend to lie near a smooth curve.

Pearson's r—Correlation Coefficient

The degree of association between two sets of data is measured by a correlation coefficient, represented by r. It is also called Pearson's correlation coefficient and measures linear association between two variables. If a curved line is needed to state the relationship, more complicated measures of correlation should be used.

The correlation coefficient is measured on a scale that varies from + 1 to – 1.

- 1 is a perfect positive correlation.
- 0 is no correlation (the values don't seem to be linked at all).
- –1 is a perfect negative correlation.

Pearson's coefficient, r, is denoted by:

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 (\Sigma y)^2]}}$$

Where,

N = Number of Values or Elements

X = First Score

Y = Second Score

ΣXY = Sum of the Product of First and Second Scores

ΣX = Sum of First Scores

ΣY = Sum of Second Scores

ΣX2 = Sum of Square of First Scores

ΣY2 = Sum of Square of Second Scores

Following are the guidelines given for interpreting the Pearson's coefficient 'r':

Coefficient, *r*		
Strength of Association	**Positive**	**Negative**
Small	.1 to .3	–0.1 to –0.3
Medium	.3 to .5	–0.3 to –0.5
Large	.5 to 1.0	–0.5 to –1.0

Note that the strength of the association of the variables depends on what you are measuring and sample sizes.

Example 1: The age and income of five people are given below. Calculate the Pearson coefficient. What does it depict?

Age (x)	Income (y)
20	2000
30	40000
40	49000
50	61000
60	75000

Solution: To calculate the coefficient, we need to calculate the following values.

x	y	xy	x^2	y^2
20	2000	40000	400	4000000
30	4500	135000	900	20250000
40	5700	228000	1600	32490000
50	6800	340000	2500	46240000
60	8000	480000	3600	64000000
$\sum x=200$	$\sum y=27000$	$\sum xy=1223000$	$\sum x^2=9000$	$\sum y^2=166980000$

Putting the values in the formula,

$$r = \frac{5\,(1223000) - (200)\,(27000)}{\sqrt{[(5)\,9000 - (200)^2]\,[(5)\,(166980000) - (27000)^2}}$$

$$= \frac{715000}{727667.5065}$$

$$= 0.98$$

0.98 represents a positive strong relationship between the two variables. As the age of a person increases, the person's income also goes up.

Example 2: Amit is an idol student good in both academics and sports. However, after some time, he reduced his sports activity and observed that he scored less marks in his test also. To investigate this hypothesis, he noted how he scored in his tests, based on how many hours he played any sport before appearing in the school tests. He gathered this data to check the correlation between number of hours of his sports activity and his tests scores. He thus, calculated the Pearson Correlation Coefficient = 0.95. Explain his observation.

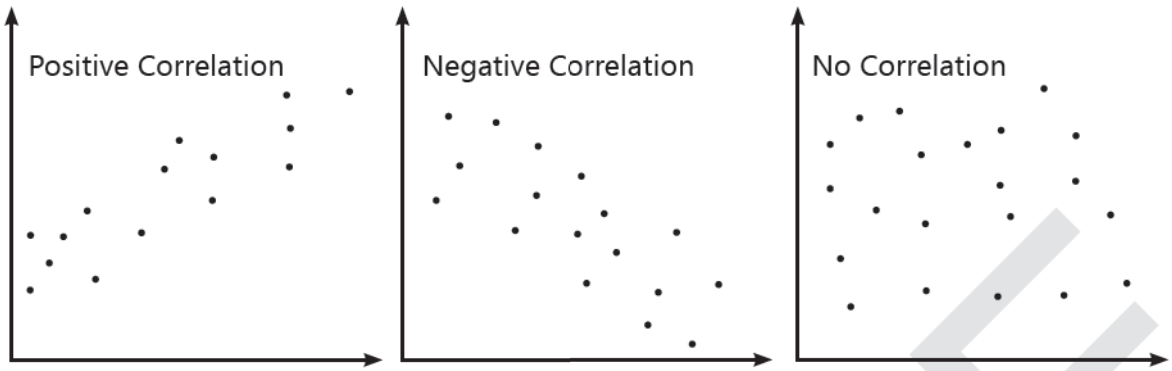
Solution: 0.95 shows a positive and strong strength of association between the two variables. This means that Amit scored better marks if he continued his sports activities. If Amit reduced his playing hours, the marks he scored also reduced.

Assumptions

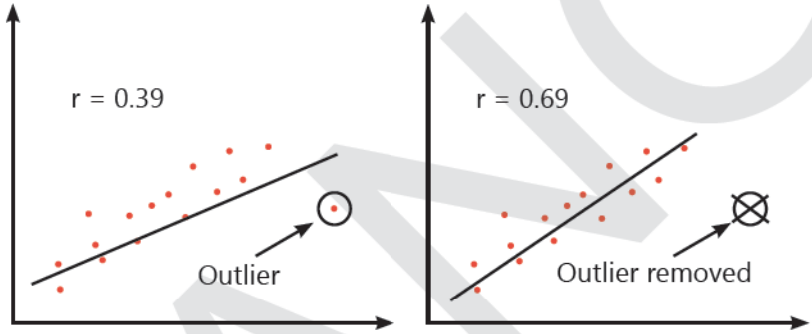
There are four assumptions for Pearson's correlation coefficient which are as follows. If any of these four requirements are not met, analysis of data using Pearson's correlation coefficient might not yield a valid result:

1. The data type of the two variables should be continuous. Examples of such continuous variables include height (measured in feet and inches), temperature (measured in °C), income (measured in INR), study time (measured in hours), intelligence (measured through IQ score), exam performance (measured from 0 to 100), sales (measured in number of transactions every month), etc.
2. There must be a linear relationship between the two variables. Create a scatterplot by plotting the two variables against each other. The scatterplot can then be used to check for linearity.

The scatterplot may look something like one of the following:



3. In statistics, outliers are data points that are significantly different from other observations. Outliers may be due to measurement irregularity or may indicate experimental error; the latter are sometimes excluded from the data set. Outliers can cause serious problems in statistical analysis. The data should not have any significant outliers. Outliers are single data points within your dataset that do not follow the usual pattern. The following scatterplots highlight the potential impact of outliers:



Outliers can have a great impact on the line of best fit and the Pearson correlation coefficient, leading to very difficult inferences regarding the data. Therefore, it is best to have no outliers or keep them to a minimum.

4. The variables should be normally distributed (approximately).

Correlation is not Causation

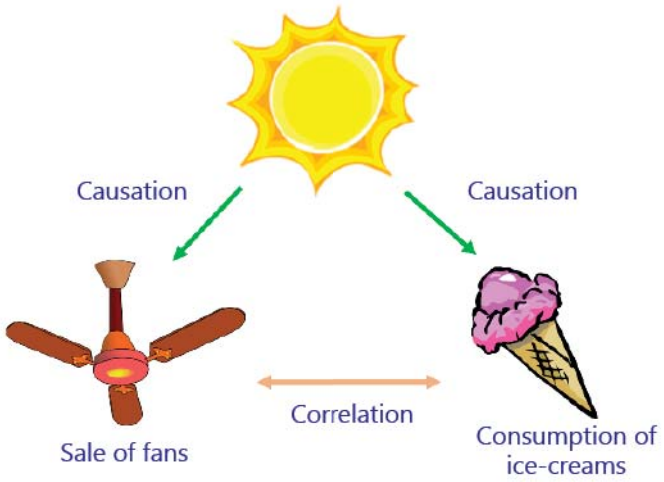
The **correlation** is a statistical method that indicates whether a pair of variables has a linear relationship and will change together. It does not state the reasons for the relationship, but it tells that a relationship exists.

Causation shows that an event is the direct result of the occurrence of another event, i.e. a causal relationship exists between the two events. This is also called **cause and effect**. For example, a speeding car leads to an accident. The accident is due to causation.

Causation takes a step ahead of correlation. It states that any change in the value of one variable will definitely cause a change in the value of the second variable. This means that one variable makes the other happen. This is also called **cause and effect**.

In statistics, the phrase "correlation does not imply causation" means that the relationship between two variables cannot be reasonably deduced based solely on their observed association.

- "Correlation is not causation" means that if two things are related, does not, necessarily mean that one thing leads to the other.
- For example, just because Indians tend to eat more in cold weather and less in hot weather does not mean that cold weather leads to crazy shopping for eatables.

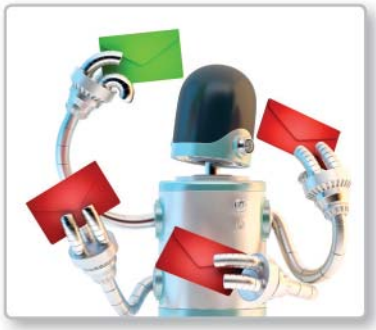


- Another example is that, due to less RAM, our mobile phone freezes. This means no playing games or messaging through the phone.

Classification

We use classification every day, classifying vegetables as 'good to eat' or 'rotten', classifying cats as per their breed or even classifying files as important or not important.

One more common example of classification is to identify spam emails. To create an email spam filter program, a developer can train a machine learning algorithm with a set of spam-like emails labelled "spam" and normal emails labelled "not spam".



The reason behind this is to train an algorithm that can identify spam emails from a set of emails and filter out non spam emails.

In order to understand "Classification", let us revise the concept of "Supervised Learning", because classification is type of supervised learning.

Supervised learning is a subcategory of Machine Learning and Artificial Intelligence. It is also known as supervised machine learning. In supervised learning, we train the machine using well-labelled data, which means the data is already marked with the correct answer. After that, the machine is given a new dataset (training data) so that the supervised learning algorithm can analyse this data and produce the correct outcome.

Let us take an example of vegetables. Now, you want to create a machine that can identify the vegetables one by one.

To do so, the first step is to train the machine with all different vegetables one by one, which may be done as follows:

- If shape of a vegetable is round with a depression at top and is red in colour, then it will be labelled as tomato.
- If the shape of a vegetable is long and conical, although cylindrical and nearly spherical, and is orange/red in colour, then it will be labelled as carrot.
- If the shape of a vegetable is long finger-like, has a small tip at the tapering end, and is green in colour, then it will be labelled as a lady finger.

Now, if you show a new vegetable in front of the machine and ask the machine to identify it, since the machine has already been trained from previous data, it will use the learned data wisely to classify the vegetable based on its shape and colour and would confirm the vegetable.

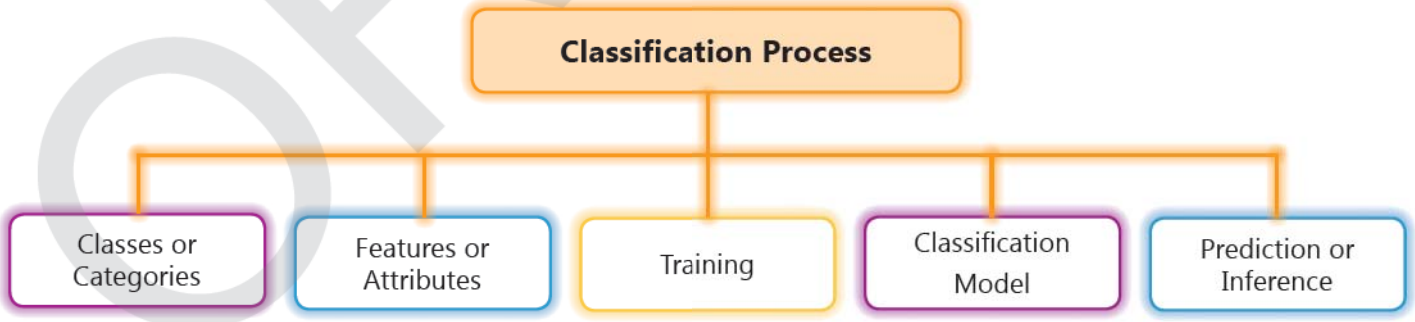
More examples of classification problems include:

- Given a handwritten character, classify it as one of the known characters.
- Given recent user behaviour, classify it as churn or not.

In Artificial Intelligence, classification is the process of labelling a set of data (structured or unstructured) into different classes or groups where we can assign a label to each class. For example, cities in India have different coloured dustbins for different types of waste: green coloured dustbins for biodegradable waste, blue dustbins for non-biodegradable or plastic waste, yellow dustbins for paper waste, and red dustbins for metallic waste. Hence, we classify the waste into four different categories while also labelling each category.

How Classification Works?

In machine learning, classification involves sorting data into specific groups or classes based on their features.



Here's an overview of the process:

- **Classes or Categories:** Data is organised into different classes or categories, each representing a distinct outcome. For example, a binary classification scenario has two classes: positive and negative.
- **Features or Attributes:** Each data instance is characterised by features or attributes that provide information about it. These features are essential for the classification model to distinguish between different classes. For example, in email classification, features might include words in the email, sender information, and the email subject.

- **Training Data:** The classification model is trained on a dataset known as training data. This dataset contains labelled examples, where each data instance is paired with a class label. The model learns the relationship between features and class labels from this data.
- **Classification Model:** An algorithm or technique is used to create the classification model. The model learns from the training data to predict class labels for new, unseen data instances, generalising patterns and relationships from the training data to make accurate predictions.
- **Prediction or Inference:** After training, the classification model is used to predict class labels for new data instances. This process, known as prediction or inference, that uses the learned patterns and relationships from the training data.

Ai Task

Form a group of 5 students. Each group should think and come up with one use case from the classroom environment or their home/society, where they would like to apply classification algorithm to solve the problem.

Real Life Applications of Classification

In classification, our data is categorised into a preferred and distinct number of classes while assigning label to each class. Applications of classification in real life include:

- Speech recognition
- Handwriting recognition
- Face tagging as done by Facebook
- Detecting fraudulent transactions in banks
- Predicting whether advertisements on a website to be clicked or not
- Product classification
- Document classification

Digital Literacy

Video Session

- Scan the QR code or visit the following link to watch the video: Classification in Machine Learning
  https://www.youtube.com/watch?v=xG-E--Ak5jg
- Scan the QR code or visit the following link to watch the video: Difference between Classification and Regression
  https://www.youtube.com/watch?v=9rRLo4nFc3A

After watching both the videos, answer the following questions:

1. Write any two differences between Classification and Regression.

2. What is classification model according to the video?

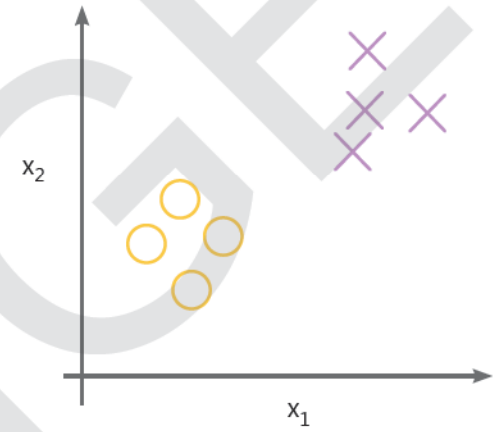
Brainy Fact

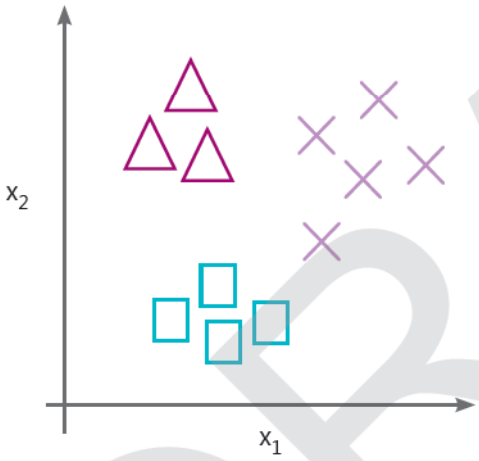
One of the first algorithms used for machine learning was the Naive Bayes classifier. Spam filtering systems used Naive Bayes till 2010.

Types of Classification

Classification is a supervised learning concept which groups a set of data into classes. It is mainly of four types, which are as follows:

- **Binary classification:** It refers to classification problems/tasks that have only two class labels. For example, email spam detection (spam or not), churn prediction (whether customers will stop doing business on a particular website or he will continue). In general, binary classification tasks involve 2 labels—**normal** and **abnormal**. For example, "no-spam" is a normal condition and "spam" is an abnormal condition. Another example is that "tumour not detected" is a normal state and "tumour detected" is an abnormal condition.

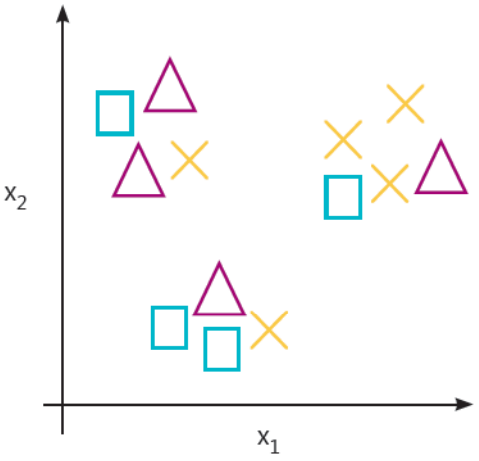


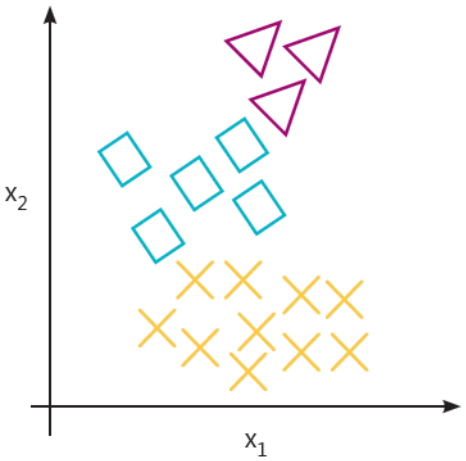


- **Multiclass classification:** It implies those classification tasks that have more than two class labels. Each entity is assigned to one class without any overlap. For example, face classification, animal species classification, optical character recognition. In contrast to binary classification, multiclass classification does not have the concept of normal and abnormal classes. Instead, the examples are classified as belonging to one of the several known classes.

  The number of classes may be very large in some problems. For example, a model may tag a photo as belonging to one among thousands of faces in a face recognition system. Text translation models are also a special type of multiclass classification.

- **Multi-label classification:** Multi-label classification is used when a situation might belong to more than one class at the same time. This implies that for a given input, the output may contain a collection of labels instead of a single one. For example, document classification (where a document may be classified into various groups at the same time, such as "science" and "technology"), and object detection in images etc.





- **Imbalanced classification:** Imbalanced classification involves tasks associated with classification in which the distribution of classes in the dataset is strongly distorted, implying that a single class (the majority class) surpasses the other(s). This could lead to problems throughout training since the model may get biased against the majority class while doing negatively on the minority class.

K-Nearest Neighbor(KNN) Algorithm

K-Nearest Neighbor (KNN) is one of the most basic yet essential classification algorithms in machine learning. It is a simple, straightforward, and adaptable machine learning method. The KNN algorithm is a classifier using supervised learning and non-parametric learning (makes no assumptions about the original data distribution) that employs proximity or closeness to classify or forecast the arrangement of a single data point.

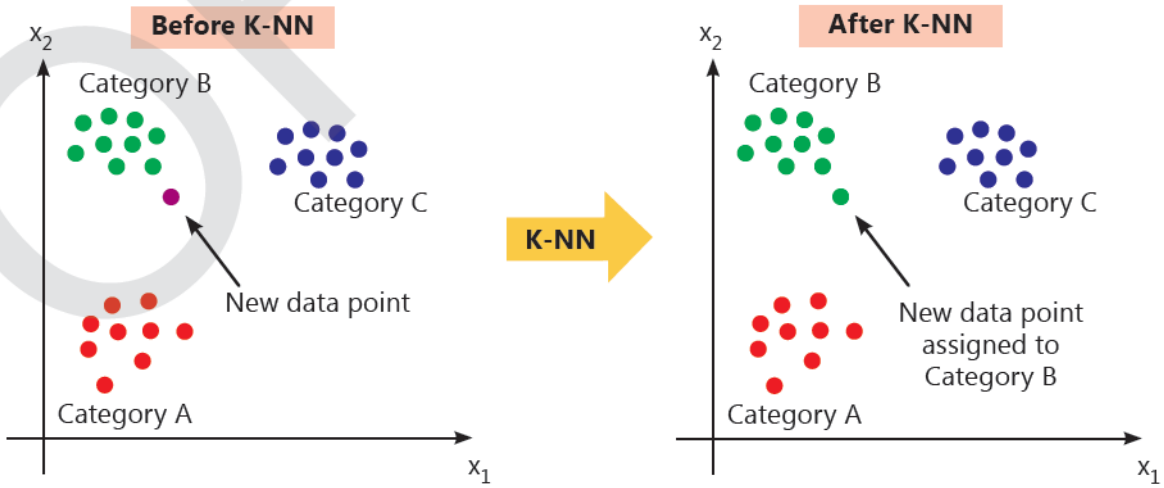
KNN is most beneficial where labelled data is prohibitively expensive or difficult to gather and it can perform well in a wide range of forecasting situations. It has a wide range of applications, including handwriting detection, picture recognition, pattern recognition and video recognition.

While the KNN technique can be applied to both regression and classification problems, it is most commonly employed for classification, with the assumption that similar data points are possible to find close proximity.

Why is KNN Algorithm Required?

Assume you have a collection of coloured balls divided into three colours: red, blue, and green. If you purchase a new ball but do not specify its colour, you will be unable to determine which colour group it belongs to. In such cases, when the decision limits are unclear or the dataset lacks a well-defined structure, the KNN algorithm is applied.

The K-Nearest Neighbor (KNN) method determines the colour of the ball based on proximity of new ball or datapoint. If the new ball is close to the red colour group, it is labelled as red; if it close to the blue colour group, it is labelled as blue; and if it close to the green colour group, it is labelled as green.



Steps Involved in KNN

To better understand the working of KNN algorithm, apply the following steps when using it:

Step 1: Load both the training and test datasets.

Step 2: Select the number of nearest data points (K), which can be any integer.

Step 3: Choose a distance metric, such as Euclidean or Manhattan distance, to determine the closeness among data points.

Step 4: Arrange the dataset in ascending order based on the distance values and determine the K-Nearest Neighbors.

Step 5: Determine the total amount of data points in every category among these K neighbors.

Step 6: Assign the new data point to the category having the highest number across its immediate neighbors.

Step 7: The model is prepared and ready for usage.

Applications of KNN

Some applications of KNN are as follows:

- **Image recognition:** KNN may be used to categorise photographs depending on their attributes, such as pixel values and colour, etc. KNN may compare the attributes of a picture with those of labelled images in the set used for training and classify the majority of its K-Nearest Neighbors.
- **Spam detection:** KNN can identify spam emails through the comparison of new emails to a database containing both spam and non-spam emails.
- **Medical diagnosis:** KNN can forecast diseases using patient information such as symptoms, medical history, and medical test findings. Patient information can be expressed as feature vectors, which KNN can use to produce forecasts by comparing them to those individuals with recognised diagnosis.
- **Financial forecasting:** KNN may be employed to anticipate stock prices and market movements using the past information. KNN can discover trends in previous data and forecast future market behaviour.
- **Anomaly detection:** Anomaly detection uses KNN to detect data points which are substantially distinct from the remaining portion of the data. It identifies points as anomalies when they are significantly different from their K-Nearest Neighbors.
- **Recommendation systems:** KNN can enhance recommendation systems by identifying comparable people or things.
- **Customer segmentation:** KNN may divide customers into groups according to purchase habits, demographic information, or other characteristics, allowing for more focused marketing campaigns.

Advantages of KNN

Some advantages of KNN are as follows:

- KNN modeling doesn't have a training period because the data itself serves as the model for future predictions. This makes it very efficient for quickly using the available data.
- KNN is also very easy to implement. The main task is to calculate the distance between different points based on their features. This can be done easily using distance formulas like Euclidean or Manhattan.
- Since there's no training period, new data can be added at any time without affecting the model.

Limitations of KNN

Some disadvantages of KNN are as follows:

- KNN is not good for large datasets because calculating distances for every data instance is very time-consuming.

- It doesn't perform well with high-dimensional data, as calculating distances for many dimensions is complex.
- KNN is sensitive to noisy and missing data.
- All data dimensions must be properly scaled (normalised and standardised).

For Advanced Learners

The Breast Cancer dataset from the sklearn library is a well-known dataset used for binary classification tasks. It contains data on breast cancer cases collected by Dr. William H. Wolberg at the University of Wisconsin Hospitals. This dataset includes features that describe the characteristics of the cell nuclei present in the breast cancer biopsies. The dataset consists of 569 rows with 30 numerical features. Each row corresponds to a biopsy sample, and each feature represents a specific characteristic of the cell nuclei like mean radius, mean texture, mean perimeter, mean area and so on. The target variable is binary, indicating whether the cancer is malignant (1) or benign (0).

Program 2: To demonstrate the use of KNN in Classification using Python

```
# Import necessary libraries
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.datasets import load_breast_cancer
from sklearn.metrics import accuracy_score

# Load the breast cancer dataset
data = load_breast_cancer()
X, y = data.data, data.target

# Convert to DataFrame for better visualization
df = pd.DataFrame(data.data, columns=data.feature_names)
df['target'] = data.target

# Display the first 10 rows of the dataset
print("First 5 rows of the dataset:")
print(df.head(5))

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Create a KNN classifier with k=3
knn = KNeighborsClassifier(n_neighbors=3)

# Fit the classifier to the training data
knn.fit(X_train, y_train)

# Predict the labels for the test set
y_pred = knn.predict(X_test)

# Calculate the accuracy of the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy * 100:.2f}%')
```

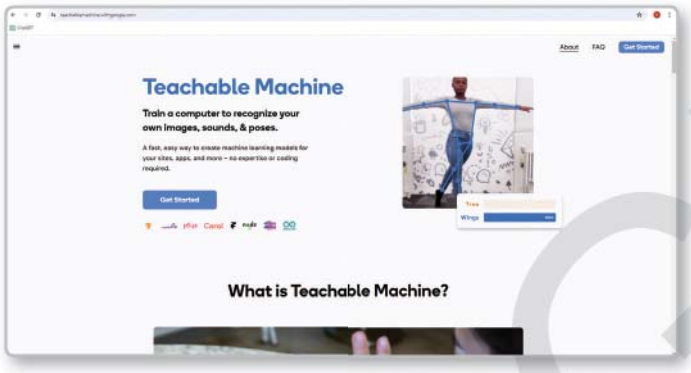
Output:

```
First 5 rows of the dataset:
      mean radius  mean texture  mean perimeter  mean area  mean smoothness  \
0           17.99         10.38          122.80     1001.0          0.11840
1           20.57         17.77          132.90     1326.0          0.08474
2           19.69         21.25          130.00     1203.0          0.10960
3           11.42         20.38           77.58      386.1          0.14250
4           20.29         14.34          135.10     1297.0          0.10030

      mean compactness  mean concavity  mean concave points  mean symmetry  \
0              0.27760          0.3001              0.14710         0.2419
1              0.07864          0.0869              0.07017         0.1812
2              0.15990          0.1974              0.12790         0.2069
3              0.28390          0.2414              0.10520         0.2597
4              0.13280          0.1980              0.10430         0.1809

      mean fractal dimension  ...  worst texture  worst perimeter  worst area  \
0                    0.07871  ...          17.33           184.60      2019.0
1                    0.05667  ...          23.41           158.80      1956.0
2                    0.05999  ...          25.53           152.50      1709.0
3                    0.09744  ...          26.50            98.87       567.7
4                    0.05883  ...          16.67           152.20      1575.0

      worst smoothness  worst compactness  worst concavity  worst concave points  \
0               0.1622             0.6656           0.7119                0.2654
1               0.1238             0.1866           0.2416                0.1860
2               0.1444             0.4245           0.4504                0.2430
3               0.2098             0.8663           0.6869                0.2575
4               0.1374             0.2050           0.4000                0.1625

      worst symmetry  worst fractal dimension  target
0             0.4601                  0.11890       0
1             0.2750                  0.08902       0
2             0.3613                  0.08758       0
3             0.6638                  0.17300       0
4             0.2364                  0.07678       0

[5 rows x 31 columns]
Accuracy: 94.15%
```

In the above program, the dataset is split into training and testing sets using train_test_split. A KNN classifier is created with k=3 neighbors. This classifier is trained using the training data. The trained model is then used to predict labels for the test set. Finally, the accuracy of the model is calculated and printed.

Ai Task

Scan the QR code or visit the following link https://teachablemachine.withgoogle.com/ to use Google's Teachable Machine to create a classification model to identify different electronic gadgets like smartphones, tablets, remote controls, headphones, USB drives etc.



Ai Reboot

1. Fill in the blank
   In Machine Learning, classification involves sorting data into specific groups or classes based on their _____.
2. State True or False
   KNN can be applied to both regression and classification problems. _____
3. What role do features or attributes play in classification?
   _____
   _____
4. List any 2 types of classification.
   _____
   _____
5. List any 2 applications of KNN algorithm.
   _____
   _____

Unsupervised Learning - Clustering

Suppose you plan to go on a holiday to Amritsar, Punjab. You plan to visit around 20 tourist spots in Amritsar in three days. How will you do this? Write down your plan.

You will most probably group the tourist spots into say 3 – 5 groups. The ones closer to each other like the Golden Temple and Jallianwala Bagh will go in one group, while Wagah Border which is further away will be included in another group. You do this, so that you can park your car at one spot/central location and walk around visiting all the tourist spots in your group/cluster.

The above can be classified as a clustering problem. Clustering, or cluster analysis, is the task of grouping a data set into a set of similar items. It is an unsupervised learning algorithm. As an important analysis method in machine learning, clustering is used to identify patterns and structures in unlabelled data sets. Clustering algorithms are used in many fields, as given below.

Applications of Clustering

There are various applications of clustering. Let us learn about some of them.

Fake News Detection

The algorithm uses the content of fake news articles, checks the words used, and then groups them into clusters. These clusters help the algorithm determine which is real news and which is fake news. If you see a high percentage of certain terms in the article, the material is more likely to be fake news.

Marketing and Sales

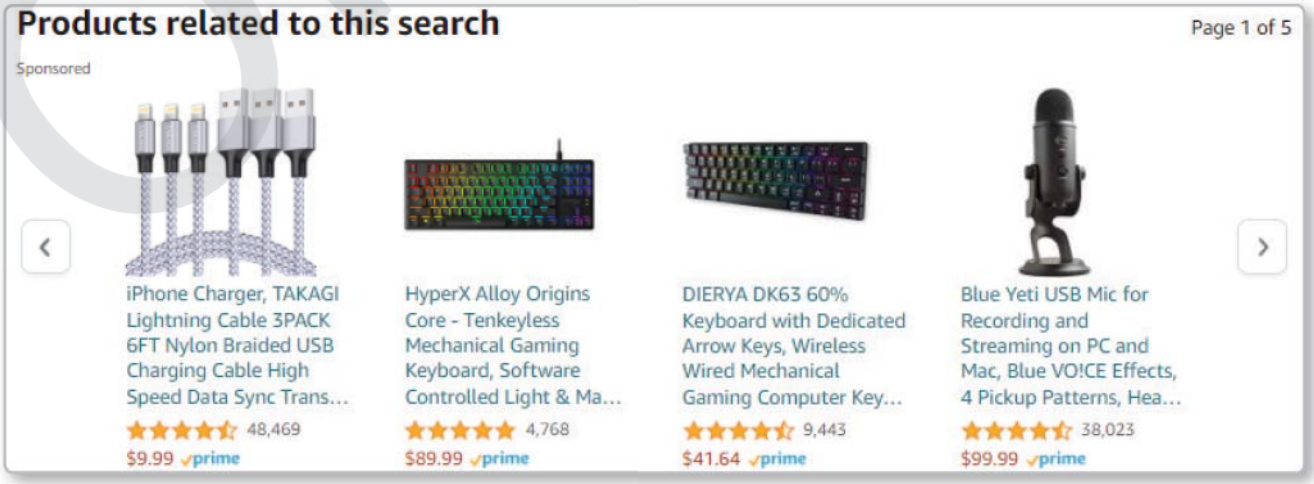
If you are a CEO of a company, targeting the right candidates is critical. Clustering algorithms group people with similar characteristics who are most likely to sell your product or service. Once these groups have been identified, target your messages around them to increase the likelihood of sales.

Document Classification/Organisation

The algorithm views the text and groups/clusters it into different topics. This technique allows to quickly group and organise similar documents using the characteristics given in the paragraph.

Recommendation Systems

Recommendation systems are widely used by Amazon, Netflix, Flipkart, etc. to provide automated and personalised recommendations for products, services and information. The technology behind the recommendation engines is called collaborative filtering. A cluster is formed based on the preferences of customers. Customers within each cluster get recommendations computed at the cluster level.



Medical Imaging Analysis

Clustering is used to match patterns in the images and identify cancerous datasets. A mix of both cancerous and non-cancerous datasets are analysed by the clustering algorithms to understand the different characteristics present in the dataset, producing resultant clusters.

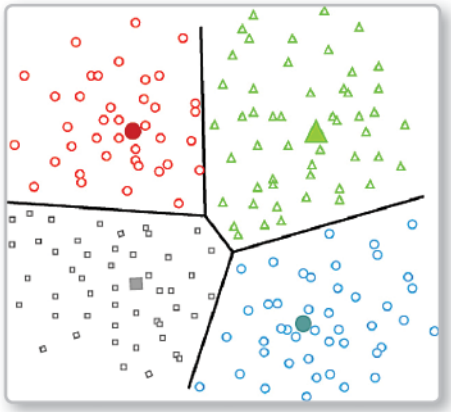
How Clustering Works?

In order to cluster the data, the following steps are conducted:

1. **Data preparation:** Data preparation means including effective data features for the clustering algorithm. The data set must include descriptive features or any new features based on the original set that will be generated, in the input dataset.
2. **Creating similarity metric:** The algorithm tries to understand how similar the pairs of samples are. You quantify the similarity between the samples by creating a similarity metric. This requires clear understanding of your data and how to derive similarity from the data features. For example, consider pin codes of an Indian state. If the difference between two pin codes is small, this represents that the two regions denoted by the pin codes are close to each other and have a higher similarity. When you can quantify the metric manually, it is called '**manual similarity measure**'.
3. **Run the clustering algorithm:** A clustering algorithm uses the similarity metric developed in step 2 to cluster data. Clustering algorithms are able to handle processing of large datasets efficiently. However, they do need to compute the similarity between all pairs.
4. **Result interpretation:** As clustering is unsupervised, the interpretation of results is crucial and can be handled by a human expert. The results are verified against expectations and if improvement is required, the above steps are repeated.

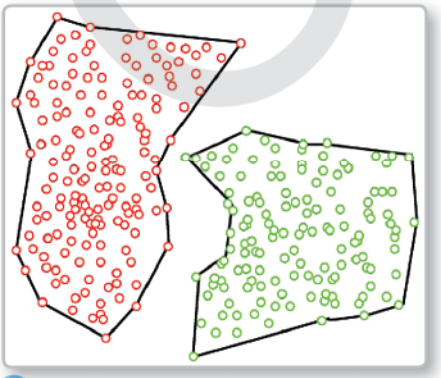
Types of Clustering

Clustering algorithms are quite popular. Let us learn about some of them.



Centroid-based Clustering

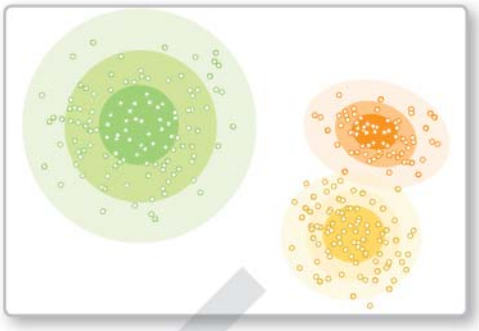
Centroid-based clustering arranges the data into non-hierarchical clusters. K-means clustering is the most popular centroid-based clustering algorithm. Centroid-based algorithms are efficient but easily affected by the initial conditions and outliers. This type of clustering is also called **Partitioning Clustering**.



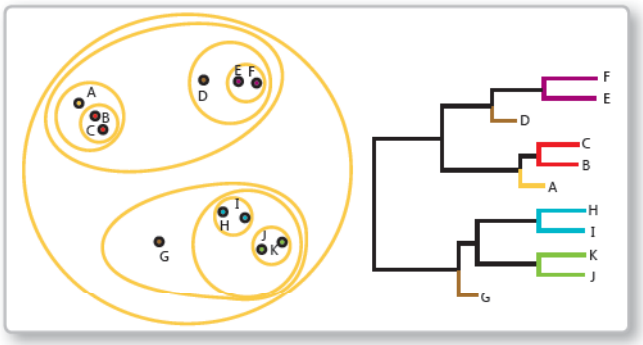
Density-based Clustering

Density-based clustering groups high density areas into clusters. Hence, arbitrary-shaped distributions occur so that dense areas can be connected. The data points in the separating regions of low density are considered outliers and not assigned to clusters.

Distribution-based Clustering



Distribution-based clustering is a clustering model in which we try to fit the data on the probability that it can belong to the same distribution. The grouping done may be normal or gaussian. **Gaussian distribution** is more popular where there are fixed number of distributions. The data is fitted in such a way that the distribution of data gets maximised.



Hierarchical Clustering

Hierarchical clustering builds a tree of clusters. The aim of the algorithm is to produce a tiered series of nested clusters. Each cluster is different from every other cluster, and the objects within each cluster are mostly similar to each other.

K-Means Clustering

Out of the various clustering techniques mentioned above, the simplest and very widely used clustering algorithm is "**centroid-based clustering using K-means**".

A **centroid** is an imaginary or real location denoting the centre of the cluster. The K-means algorithm identifies K number of centroids, and then assigns every data point to the nearest cluster, while trying to keep the centroids as small as possible.

The algorithm has the following steps:

Step 1: Decide the number of clusters (k).

Step 2: Select k random points from the data as centroids.

Step 3: Group all the points to the nearest centroid.

Step 4: Calculate the centroid of newly formed clusters.

Step 5: Repeat steps 3 and 4.

It is a repetitive process. It will keep on executing until there is no change in the centroids of newly formed clusters or points remain in the same cluster or the maximum number of iterations are reached.

Digital Literacy

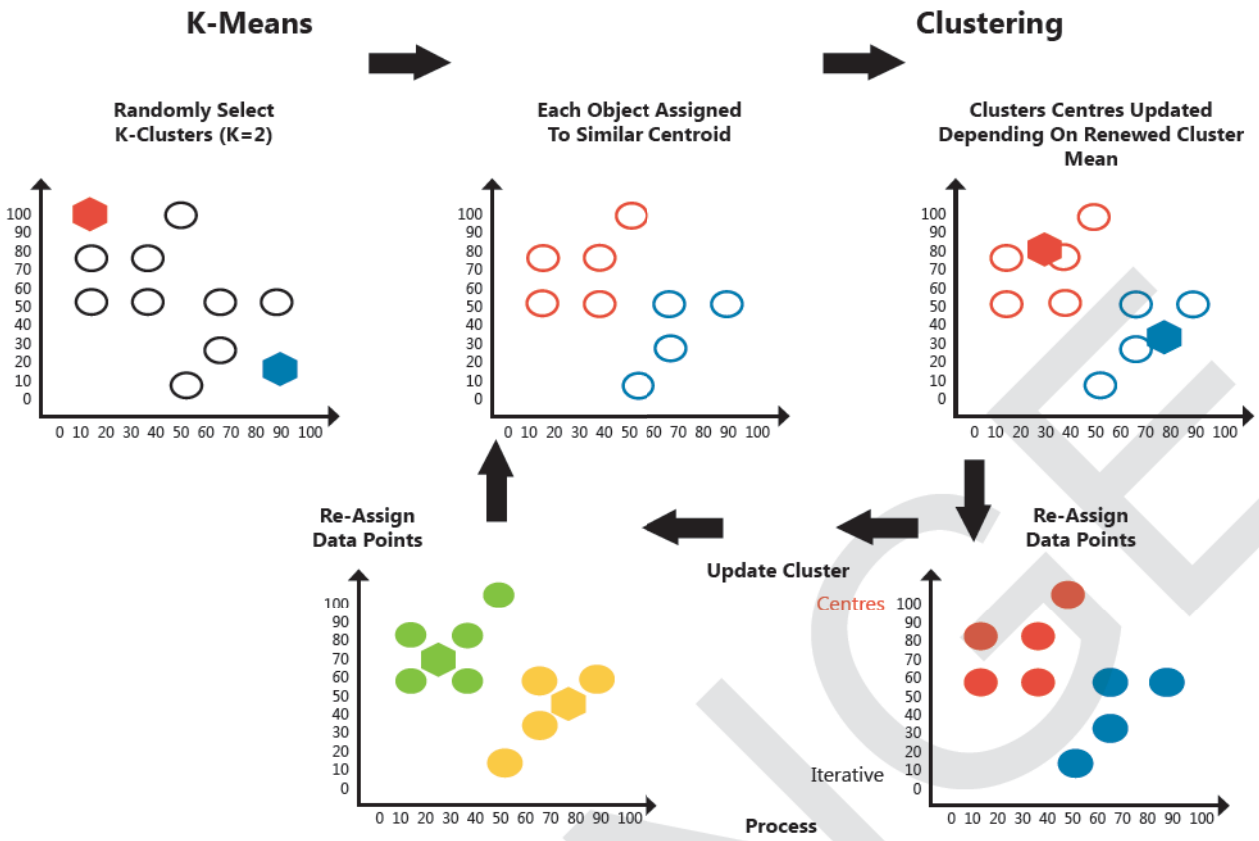
Video Session

Scan the QR code or visit the following link to watch the video: StatQuest: K-means Clustering

https://www.youtube.com/watch?v=4b5d3muPQmA

After watching the video, answer the following question:

What do you mean by K-means clustering according to the video?



Advantages of K-Means Clustering

Some of the advantages of K-Means Clustering are:

- Easy to implement.
- Can handle large data sets.
- Can give initial positions to centroids (randomly).
- Easily adapts to new data.
- Can easily adapt to clusters of different shapes and sizes, like elliptical clusters.

Disadvantages of K-Means Clustering

Some of the disadvantages of K-Means Clustering are:

- K has to be chosen manually and it is not an easy process.
- The algorithm is dependent on initial values.
- Outliers greatly affect the clustering process.
- The algorithm has trouble grouping data where clusters are of fluctuating sizes and density.

Why is Clustering Unsupervised?

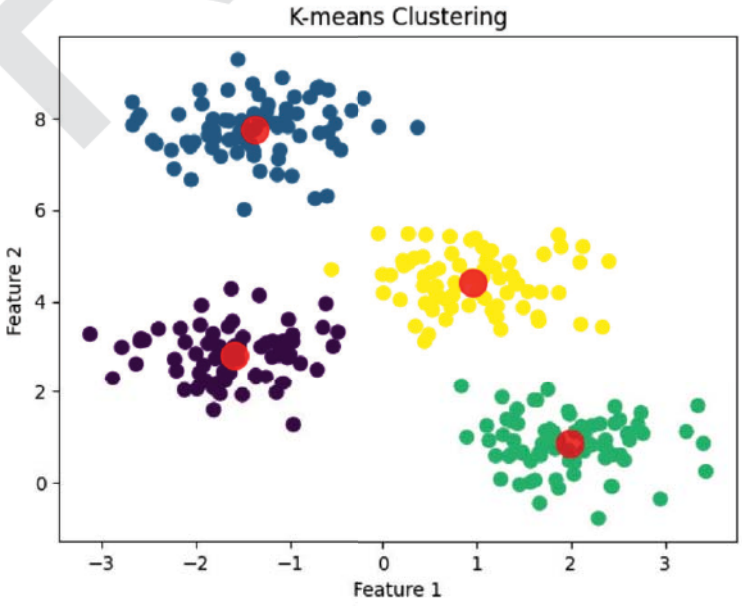
Clustering is an unsupervised machine learning technique that automatically divides the data into clusters or groups of similar elements. The algorithm does this without any knowledge of how the groups should look in advance. So, clustering is rather used for the discovery of knowledge rather than for prediction. It provides an idea of natural groupings that are within data.

Without advanced knowledge of what a cluster includes, how can a computer know where a group begins or ends? The answer is simple. Clustering is driven by the principle that objects within a group should be very similar to each other, but very different from the objects outside. The similarity function can vary across different applications, but the basic idea is always the same—group the data so that the related elements are placed together.

For Advanced Learners

Program 3: Write a program to represent K Means Clustering using Python.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make_blobs
from sklearn.cluster import KMeans

# Generate synthetic data
X, _ = make_blobs(n_samples=300, centers=4, cluster_std=0.60, random_state=0)

# Apply K-means clustering
kmeans = KMeans(n_clusters=4)
kmeans.fit(X)
y_kmeans = kmeans.predict(X)

# Plot the data points and centroids
plt.scatter(X[:, 0], X[:, 1], c=y_kmeans, s=50, cmap='viridis')
centers = kmeans.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], c='red', s=200, alpha=0.75)
plt.title('K-means Clustering')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.show()
```



This program does the following:

1. Generates synthetic data using make_blobs from sklearn.datasets.
2. Applies K-means clustering with n_clusters=4.
3. Plots the data points colored by their cluster assignments and shows the centroids as red circles.

   You can adjust the parameters like the number of clusters, standard deviation, and number of samples in make_blobs to observe different clustering scenarios.

Ai Reboot

1. Which type of Machine Learning uses labelled data?

2. Which two types of Machine Learning do not require supervision?

3. Which ML is reward-based?

4. Which type of ML is used by Netflix's recommender systems?

5. A robot learns to walk on its own. Which type of ML is this?

Ai Task

Experiential Learning

Check out the website https://lipsync.withyoutube.com/

An AI-powered challenge that rates how closely your lip-syncing matches the song! This experiment uses Google's AI technology TensorFlow.js. The TensorFlow model provides a real-time high-density estimate of key points of your facial expression using only a webcam. Through machine learning in your browser, the key points around the mouth and lips are analysed to estimate how well you synchrosise to the lyrics of the song!

At a Glance

- Machine learning is an application of Artificial Intelligence (AI) that enables systems to learn and improve automatically from experience without the need for explicit programming.
- Machine learning can be broadly categorised into three main types based on the learning approach and input data: Supervised Learning, Unsupervised Learning and Reinforcement Learning.
- In supervised learning, the algorithm learns from labelled data, where each training example is paired with a corresponding target label.

- There are two algorithms of supervised learning named regression and classification.
- Unsupervised machine learning algorithms are used when the information used to train is neither classified nor labelled.
- Clustering and Dimensionality Reduction are common tasks in unsupervised learning.
- Regression is a Machine Learning algorithm used to analyse the relationship among dependent variable (target) and independent variable (predictor). It predicts the output values based on input values.
- Regression is basically used when the dependent variable is of a continuous data type. The independent variables, on the other hand, can be of any data type—continuous, nominal/categorical etc.
- There are several types of regression analysis, random forest regression, support vector regression, decision tree regression, linear regression, polynomial regression, ridge regression, lasso regression and logistic regression.
- Classification is a supervised learning concept which groups a set of data into classes.
- Classification is mainly of four types, binary classification, multiclass classification, multi-label classification, and imbalanced classification.
- The KNN algorithm is a classifier using supervised learning and non-parametric learning (makes no assumptions about the original data distribution) that employs closeness to classify or forecast the arrangement of a single data point.
- Clustering, or cluster analysis, is the task of grouping a data set into similar items.
- Clustering is of four types, centroid-based clustering, density-based clustering, distribution- based clustering, and hierarchical clustering.
- K-Means clustering is a powerful and widely used machine learning algorithm for partitioning datasets into clusters based on feature similarity.
- The degree of association is measured by a correlation coefficient, represented by r. It is also called Pearson's correlation coefficient and measures linear association between two variables.
- Causation shows that an event is the direct result of the occurrence of another event, i.e. a causal relationship exists between the two events. This is also called cause and effect.
- The confusion matrix compares the actual target values with those predicted by the classifier. This tells us how well the classification model is performing and what kind of error it is making.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AiQuiz

A. Tick (✓) the correct option.

1. Which of the following best describes the primary capability of machine learning?
   a. It allows computer systems to acquire knowledge and improve through experience without being specifically programmed.
   b. It is a type of hardware that increases computer processing speed.
   c. It involves manually coding each function and rule that the system must follow.
   d. It is a method to directly control hardware using machine code.

2. Reinforcement learning is based on _____ .
   a. Learning from labelled data
   b. Learning through observation
   c. The trial and error method
   d. Dimensionality reduction

3. What is the goal of linear regression?
   a. To find a polynomial relationship between variables
   b. To estimate the probability of an event
   c. To fit a straight line that minimises the sum of squared differences between observed and predicted values
   d. To group data into clusters

4. Which of the following is used when a situation might belong to more than one class at the same time?
   a. Multi-label classification
   b. Binary classification
   c. Multi-class classification
   d. Imbalanced classification

5. For which one of these relationships could we use a regression analysis?
   a. Relationship between political party membership and opinion about Euthanasia.
   b. Relationship between gender and whether person has a mole
   c. Relationship between weight and height
   d. Relationship between eye colour (blue, brown, etc.) and hair colour (black, blond, etc.).

6. The correlation between two variables is given as r = 0.0. What does this mean?
   a. The best straight line through the data is horizontal.
   b. There is a perfect positive relationship between the two variables.
   c. There is a perfect negative relationship between the two variables.
   d. All of the points must fall exactly on the horizontal line.

7. Which of the following is true about residual error?
   a. lower value is better
   b. higher value is better
   c. Either a. or b.
   d. All of these

8. Which of the following is NOT typically a use case for regression analysis?
   a. Weather forecasting
   b. Finding the causal-effect relationship between variables
   c. Time series modelling
   d. Grouping data into specified categories

9. Which of the following is NOT an application of linear regression?
   a. Predicting the effect of fertiliser on crop yield
   b. Managing risks and portfolio optimisation
   c. Classifying images of handwritten digits
   d. Predicting salary based on years of experience

10. What is the main goal of clustering in machine learning?
    a. To label data points with predefined categories.
    b. To divide a dataset into groups where items in the same group are more similar to each other than to those in other groups.
    c. To reduce the dimensionality of the data.
    d. To find associations between different variables in the dataset.

B. Fill in the blanks.

1. _____ have the capacity to extract information from limited number of datasets.
2. _____ is a technique used to reduce the number of features or variables in a dataset while preserving the most important information.
3. If data points are closer to the line of best fit (less residual error), it means the correlation between the two variables is _____ .
4. _____ in statistics, are data points that are significantly different from other observations.
5. In K-means algorithms, the value of K, which represents the number of clusters, must be selected _____.
6. _____ uses KNN to detect data points which are substantially distinct from the remaining portion of the data.
7. _____ is a clustering model in which we try to fit the data on the probability that it can belong to the same distribution.
8. _____ is driven by the principle that objects within a group should be very similar to each other, but very different from the objects outside.
9. _____ uses a single independent variable for forecasting an outcome of a numerical dependent variable.
10. In _____, the algorithm learns from labelled data, where each training example is paired with a corresponding target label.

C. State whether these statements are true or false.

1. In Supervised Learning, during training, the algorithm adjusts its parameters to minimise the difference between predicted and actual labels. _____
2. No correlation means that there is no relationship between two variables. _____
3. Bias in training data can lead to skewed predictions. _____
4. Non-parametric learning makes no assumptions about the original data distribution. _____
5. Clustering is used to identify patterns and structures in unlabelled data sets. _____
6. Single linear regression demonstrates a connection between two or more independent variables and the associated variables that are dependent. _____
7. The K-means algorithm identifies 5 centroids. _____
8. Market segmentation involves dividing a broad consumer or business market into sub-groups of consumers based on some type of shared characteristics. _____
9. Clustering algorithms group people with similar characteristics who are most likely to buy your product or service. _____
10. KNN retains few pieces of training data, requiring little memory resources. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions:

1. What is Machine learning?

Ans. Machine learning is an application of Artificial Intelligence (AI) that enables systems to learn and improve automatically from experience without the need for explicit programming. It focuses on the development of computer programs that can access data and use it to learn for itself. Data is critical for machine learning to work. The more data the machine is given (assuming that the data is reliable), the more accurate is its prediction.

2. Write down any 5 key features of machine learning.

Ans. Following are the key features of machine learning:

- Machine learning interprets, analyses, and processes data to address real-world problems.
- It learns from data and enhances its performance over time.
- The technology facilitates automation and prediction based on the learned patterns.
- Machine learning is the prevailing approach in contemporary AI.
- It employs data analysis, training, and sometimes human review to refine its capabilities.

3. List any two applications of KNN.

Ans. Two applications of KNN are as follows:

- **Image recognition:** KNN may be used to categorise photographs depending on their attributes, such as pixel values and colour, etc. KNN may compare the attributes of a picture with those of labelled images in the set used for training and classify the majority of its K-Nearest Neighbors.
- **Spam detection:** KNN can identify spam through the comparison of new emails to a database containing both spam and non-spam emails.

4. List down the real life applications of linear regression.

Ans. Real life applications of linear regression include:

- Prediction of product demand
- Sales forecasting
- Analysing the effect of price change of a service
- Predict the effect of fertiliser on crop yield
- Prediction of revenue through advertisements
- Predicting salary of a person based on the number of years of experience

5. What are the four types of correlation?

Ans. Four Types of correlation are:

1. **Positive Correlation:** Positive correlation is the relationship between two variables, in which both variables have a linear relationship. As one variable increases/decreases, the second variable too increases/decreases. For example, when fuel prices increase, prices of airline tickets also increase.
2. **Negative Correlation:** Negative correlation is the relationship between two variables, where one variable increases as the second variable decreases, and vice versa. For example, more exercising leads to a decrease in body weight.
3. **No Correlation:** No correlation means that there is no relationship between two variables. If the value of a variable is changed, another variable is not affected. For example, shirt size and monthly expense, body weight and intelligence, etc.
4. **Non-linear Correlation:** A non-linear correlation is a correlation in which all the points of a scatter plot are tend to lie near a smooth curve.

6. Differentiate between correlation and regression.

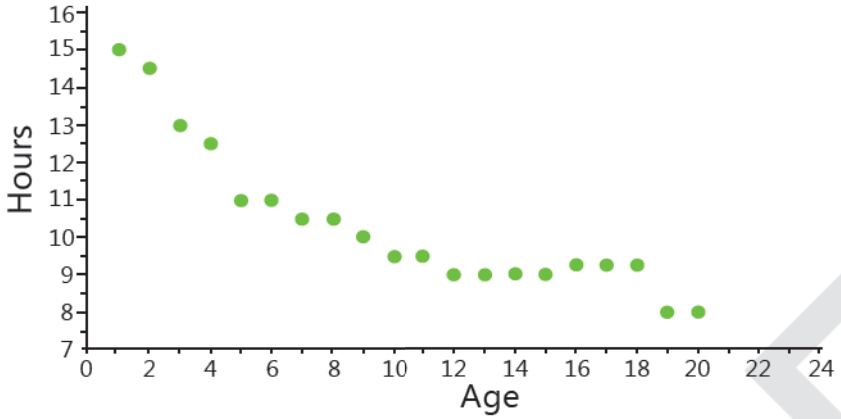
Ans. Both Correlation and regression are statistical measures used in data analysis, however they are not same. Their differences can be seen:

Correlation	Regression
It determines the strength or degree of relationship between two variables.	It determines how one variable affects another variable.
It is represented by a single value.	It is represented by a regression line.

7. What is the primary difference between classification and regression? [CBSE Handbook]

Ans. Classification predicts discrete values, while regression predicts continuous values.

8. Observe the scatter plot showing the amount of sleep needed per day by age. [CBSE Handbook]



What type of correlation is shown here?

Ans. As age increases (moving along the x-axis toward greater numbers), the amount of sleep needed decreases (y-values decreasing). This is a negative correlation. This indicates that as individuals grow older, they generally require less sleep.

9. Write down any 4 limitations of KNN.

Ans. Some limitations of KNN are as follows:

- Appropriate k selection is crucial in KNN since it has a significant impact on model accuracy, requiring careful tuning.
- KNN has issues with imbalanced datasets, that are biased towards the dominant class and generate incorrect projections for minority classes.
- KNN's evaluation stage might be highly computational because it calculates distance for every training occurrence.
- KNN retains every piece of training data, requiring large memory resources.

10. What do you understand by Centroid?

Ans. A centroid is an imaginary or real location denoting the centre of the cluster. The K-means algorithm identifies K number of centroids, and then assigns every data point to the nearest cluster, while trying to keep the centroids as small as possible.

B. Long answer type questions:

1. Define the following terms and give examples from real life:
   a. Causation
   b. Outlier

Ans: a. **Causation:** Causation shows that an event is the direct result of the occurrence of another event, i.e. a causal relationship exists between the two events. This is also called cause and effect. For example, a speeding car leads to an accident. The accident is due to causation.

b. **Outlier:** In statistics, outliers are data points that are significantly different from other observations. Outliers may be due to measurement irregularity or may indicate experimental error; the latter are sometimes excluded from the data set. Outliers can cause serious problems in statistical analysis.

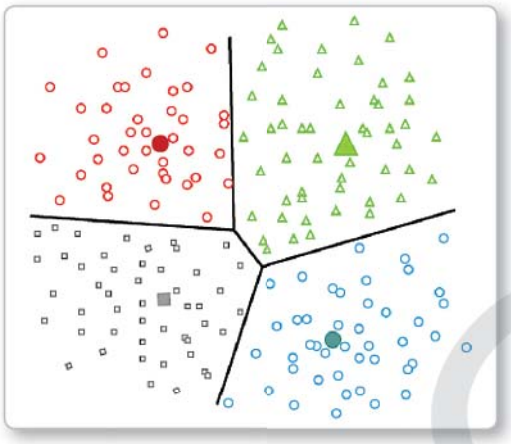
2. Explain Classification with example of supervised learning.

Ans. In Artificial Intelligence, classification is the process of labelling a set of data (structured or unstructured). into different classes or groups where we can assign a label to each class. In machine learning, a predictive classification model tries to approximately map the function from input variables to discrete output variables. The main goal is to determine which class/category the new data will belong to. For example, heart disease detection is a classification problem. There are only two classes in this case—a patient has heart disease or does not have heart disease. In this case, the classifier needs training data to understand how a given input variable is related to the class. Once the classifier is properly trained, it can be used to determine whether a particular patient has heart disease or not.

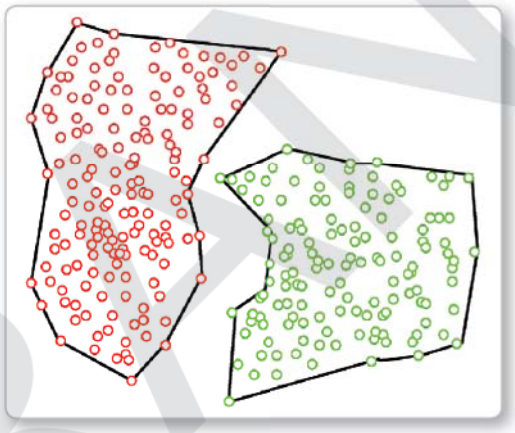
3. Discuss popular clustering algorithms.

Ans. Some popular clustering algorithms are as follows:

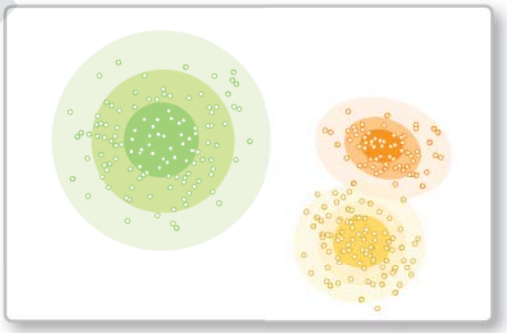
- **Centroid-based Clustering**: Centroid-based clustering arranges the data into non-hierarchical clusters. K-means clustering is the most popular centroid-based clustering algorithm. Centroid-based algorithms are efficient but easily affected by the initial conditions and outliers.



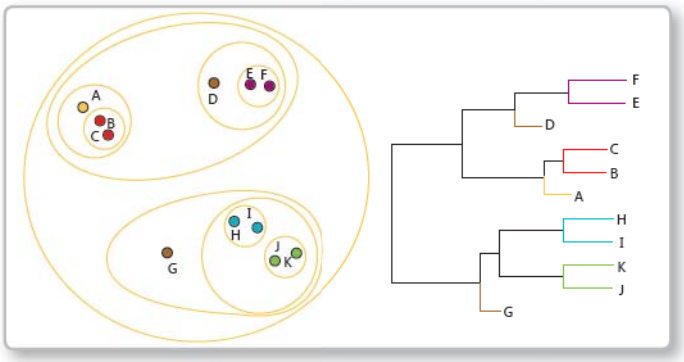
- **Density-based Clustering**: Density-based clustering groups high density areas into clusters. Hence, arbitrary-shaped distributions occur so that dense areas can be connected. The data points in the separating regions of low density are considered outliers and not assigned to clusters.



- **Distribution-based Clustering:** Distribution-based clustering is a clustering model in which we try to fit the data on the probability that it can belong to the same distribution. The grouping done may be normal or Gaussian. Gaussian distribution is more popular where there are fixed number of distributions. The data is fitted in such a way that the distribution of data gets maximised.



- **Hierarchical Clustering:** Hierarchical clustering builds a tree of clusters. The aim of the algorithm is to produce a tiered series of nested clusters. Each cluster is different from every other cluster, and the objects within each cluster are mostly similar to each other.



4. Why is clustering unsupervised?

Ans. Clustering is an unsupervised machine learning technique that automatically divides the data into clusters or groups of similar elements. The algorithm does this without any knowledge of how the groups should look in advance. So, clustering is rather used for the discovery of knowledge rather than for prediction. It provides an idea of natural groupings that are within data.

Without advanced knowledge of what a cluster includes, how can a computer know where a group begins or ends? The answer is simple. Clustering is driven by the principle that objects within a group should be very similar to each other, but very different from the objects outside. The similarity function can vary across different applications, but the basic idea is always the same—group the data so that the related elements are placed together.

5. Why is KNN algorithm required? Explain with the help of example.

Ans. The K-Nearest Neighbor (KNN) method determines the colour of the ball based on proximity of new ball or datapoint. If the new ball is close to the red colour group, it is labelled as red; if it close to the blue colour group, it is labelled as blue; and if it close to the green colour group, it is labelled as green.

C. Competency-based/Application-based questions:

1. During a class discussion, Sumit, the naughty geek of the school, enquired that how we can differentiate between Machine Learning and Artificial Intelligence. He was very well aware of the definitions but he was looking for concrete examples that actually depicted the connection and distinguish between the two. Would you help Ms. Geeta, the computer science teacher, present in the class to come up with the convincing examples for the same.

Ans. Machine learning (ML) is a subset of Artificial Intelligence (AI) that involves the development of algorithms that allow computers to learn from and make decisions based on the visual data. Following examples illustrates the same:

In AI, we conceptualise techniques to perform visual perception, speech recognition and decision making.

When we enter the sphere of machine learning, let's say we target on classification of images we have to develop classification software using algorithm like (Convolutional Neural Networks) CNN algorithm. This example shows ML as employer of a specific method used within the broader AI goal of enabling computers to understand and process visual information

Another Example is when AI aims to understand the nuances of human speech, we work upon algorithms like RNN (Recurrent Neural Networks), this part of working upon an algorithm is ML where as the goal of achieving the understanding of human language and every step proceeding towards it comes under AI.

Assertion and reasoning questions:

Direction: Questions 2-4 below, consist of two statements – Assertion (A) and Reasoning (R). Answer these questions selecting the appropriate option given below:

a. Both A and R are true and R is the correct explanation of A.

b. Both A and R are true but R is not the correct explanation of A.

c. A is true but R is false.

d. A is false but R is true.

2. **Assertion (A):** Clustering is widely used in various applications such as customer segmentation, image recognition, and data analysis to identify natural groupings within data without predefined labels.

   Reasoning (R): Clustering is a technique in machine learning where data points are grouped together based on their similarities.

Ans. A

3. **Assertion (A):** KNN's evaluation stage is very quick.
   Reasoning (R): KNN calculates distance for every training occurrence.

Ans. D

4. **Assertion (A):** Multiclass classification implies those classification tasks that have more than two class labels.

   Reasoning (R): Multiclass classification does not have the concept of normal and abnormal classes.

Ans. B

5. Two statements are given below. Statement 1 and Statement 2 . Examine the statements and answer the question according to the instruction given below.

   (A) Statement 1 is TRUE , Statement 2 is TRUE

   (B) Statement 1 is FALSE , Statement 2 is FALSE

   (C) Statement 1 is TRUE , Statement 2 is FALSE

   (D) Statement 1 is FALSE , Statement 2 is TRUE

   (i) **Statement 1:** KNN is more susceptible to outliers in the data relative with certain other algorithms.

   Statement 2: Whenever K-Nearest Neighbors are taken into account for prediction, the impact of a single outlier is increased.

Ans. (D)

   (ii) **Statement 1:** Data preparation means including effective data features for the clustering algorithm.

   Statement 2: The data features must stay within the limitations given by original data set and should not include new changes.

Ans. (C)

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. Which of the following is NOT a common machine learning method?

   a. Decision trees ◯ b. Neural networks ◯

   c. Support vector machines ◯ d. Polynomial regression ◯

2. Which type of learning does a Virtual Personal Assistant (VPA) like Siri or Alexa use?

   a. Supervised learning ◯ b. Unsupervised learning ◯

   c. Reinforcement learning ◯ d. All of the above ◯

3. In regression analysis, what type of events are typically forecasted?

   a. Categorical ◯ b. Continuous ◯

   c. Binary ◯ d. Discrete ◯

4. A regression between foot length (dependent variable in cm) and height (independent variable in inches) for 33 students resulted in the following equation:

$$y = 10.9 + 0.23x$$

A student in the sample was 71 inches tall. What is the predicted foot length for this student?

a. 17.57 cm ◯ b. 27.23 cm ◯

c. 29 cm ◯ d. 33 cm ◯

5. _____ is a classification problem.

a. heart disease detection ◯ b. recommended systems ◯

c. sales prediction ◯ d. None of these ◯

6. _____ is the process of grouping a data set into a set of similar items.

a. Regression ◯ b. Decision Trees ◯

c. Classification ◯ d. Clustering ◯

7. What is the goal of machine learning?

a. The goal of Machine Learning is to enable machines to learn manually using the provided data and make accurate Predictions/ Decisions. ◯

b. The goal of Machine Learning is to enable humans to learn manually using the provided data and make accurate predictions / decisions. ◯

c. The goal of Machine Learning is to enable machines to learn by themselves using the provided data and make accurate predictions / decisions. ◯

d. The goal of Machine Learning is to make humans lazy . ◯

8. Which of the following is NOT a clustering algorithm?

a. Distribution-based ◯ b. Centroid-based ◯

c. Multi-class ◯ d. Hierarchical ◯

9. What is one of the main disadvantages of linear regression?

a. It is difficult to implement ◯

b. It requires a large amount of computational resources ◯

c. It is sensitive to outliers ◯

d. It is not suitable for any real-life applications ◯

10. What is a primary disadvantage of the K-Means clustering algorithm?

a. It cannot handle large datasets. ◯

b. It can handle large data sets. ◯

c. It is heavily influenced by outliers. ◯

d. It is complex to implement. ◯

B. Fill in the blanks.

1. The more data the machine is given (assuming that the data is reliable), the more accurate is its _____ .
2. As the independent variable is adjusted, the level of the _____ variable will vary.
3. Centroid-based clustering arranges the data into _____ clusters.
4. _____ shows that an event is the direct result of the occurrence of another event.

5. A _____ is an imaginary or real location denoting the center of the cluster.
6. Groups of similar items are called _____ .
7. In the linear regression equation y = mx+b, m represents _____ .
8. 1 is a perfect positive correlation, _____ is a perfect negative correlation, _____ is no correlation.
9. _____ measures the strength or degree of relationship between two variables.
10. The data type of the two variables in linear regression should be _____ .

C. State whether these statements are true or false.

1. The independent variable is the variable under study, and it is the variable that the regression model tries to predict. _____
2. Linear Regression has its limitations, but its simplicity, interpretability, and efficiency often exceed these limitations. _____
3. Outliers do not affect the clustering process. _____
4. A scatterplot is actually a graph of a curvilinear function. _____
5. If a regression line that was calculated by least squares method is plotted on a scatterplot, all the points in the dataset should be on the line. _____
6. Centroid-based clustering arranges the data into hierarchical clusters. _____
7. Density-based clustering groups high density areas into clusters. _____
8. Hierarchical clustering builds a tree of clusters. _____
9. Recommendation systems are widely used by Amazon, Netflix, Flipkart, etc. to provide automated and personalised recommendations for products, services and information. _____
10. Email spam filter uses unsupervised learning algorithm. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions:

1. List 2 advantages of KNN algorithm.
2. Explain the conceptual difference between supervised learning and unsupervised learning.
3. What is Reinforcement learning?
4. What is line of least square errors?
5. What are the advantages and disadvantages of Linear Regression?
6. Justify "Correlation is not Causation".
7. Interpret the values of correlation coefficient when r=1, -1 and 0.
8. Differentiate between classification and clustering graphically.

B. Long answer type questions:

1. Explain any five types of Regression.
2. Define Linear Regression and list its applications.
3. State four assumption for Pearson's correlation.
4. Differentiate between Binary and Multiclass Classification.
5. List the steps for K-means clustering.

C. Competency-based/Application-based questions:

1. The diagrams used in KNN algorithm and K-means were very similar. An employee in Teeksha Global Tech was facing high ambiguity, as to when we should use either of these as both of these lead to classification. Could you guide her with some examples, which algorithm to use and when?
2. Vivek a Tech expert was taking an AI workshop in a school. After discussing various algorithms and validations, he announced "Confusion Matrix actually solves the confusion". Justify him.

Assertion and reasoning questions:

Direction: Questions 3-4, consist of two statements – Assertion (A) and Reasoning (R). Answer these questions selecting the appropriate option given below:

(a) Both A and R are true and R is the correct explanation of A.

(b) Both A and R are true but R is not the correct explanation of A.

(c) A is true but R is false.

(d) A is false but R is true.

3. **Assertion(A):** Causation states that any change in the value of one variable will definitely cause a change in the value of the second variable.

   Reasoning(R): The correlation is a statistical method that indicates whether a pair of variables has a quadratic relationship with each other.

4. **Assertion(A):** Linear Regression method finds the most accurate straight line that best describes the relationship between the dependent and the independent variables, with minimum error.

   Reasoning(R): Linear Regression has its limitations, but its simplicity, interpretability, and efficiency often exceed these limitations.

5. Two statements are given . Statement 1 and Statement 2 . Examine the statements and answer the question according to the instruction given below.

   (A) Statement 1 is TRUE , Statement 2 is TRUE

   (B) Statement 1 is FALSE , Statement 2 is FALSE

   (C) Statement 1 is TRUE , Statement 2 is FALSE

   (D) Statement 1 is FALSE , Statement 2 is TRUE

   (i) **Statement 1:** It is inefficient to train the machine on Linear Regression model.

   Statement 2: Linear Regression Model is quite prone to overfitting.

   (ii) **Statement 1:** Clustering is an unsupervised machine learning technique that automatically divides the data into clusters or groups of similar elements.

   Statement 2: The biggest advantage of the K-means algorithm is that it can cluster large data sets quite efficiently.

AI In Life

Interdisciplinary

Have you heard about DNA Sequence Classification? Find out about this topic and present your findings in class.

Ai Deep Thinking

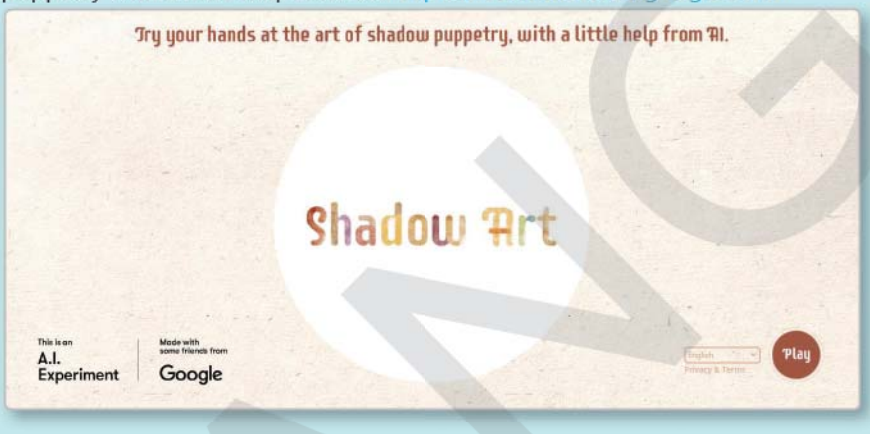
Problem solving & Logical Reasoning

Suppose you are working in a pharmaceutical company. Your team has to collect and analyse data regarding shelf life of a drug released by your company. Find out ways in which regression can help you. (Hint - Risk analysis, better decision making etc.)

Ai Lab

Experiential Learning

1. Try shadow puppetry with a little help from AI: https://shadowart.withgoogle.com



2. Select a video to watch and see how accurate the program determines your emotional engagement while you watch the video: https://demo.mr.affectiva.com/ (the website will seek permission to use your webcam).

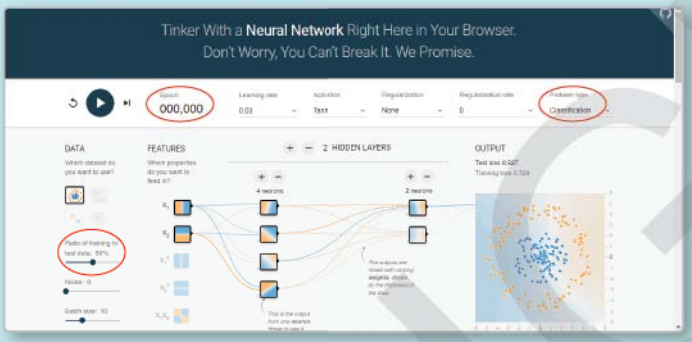
3. Visualise K-Means Clustering! Go to https://www.naftaliharris.com/blog/visualizing-k-means-clustering/ and see how clustering works in real time.

4. A shopkeeper sells pens in his shop. He sells the pens for a different cost price for six weeks. He notes the number of pens, y, sold each week. The results are shown in the table below:

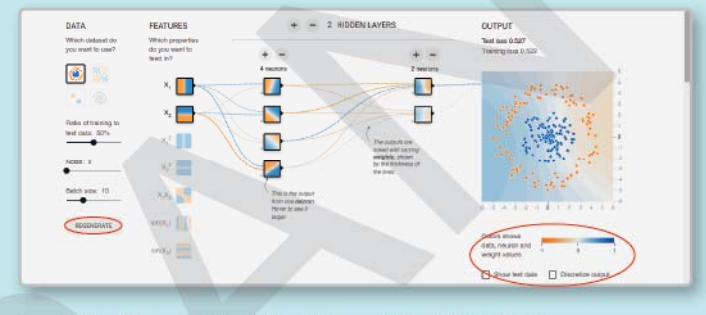
x (cost price)	15	20	25	30	35	40
y (no. of pens sold)	65	60	55	45	38	35

Plot the scatterplot for the above using MS Excel. Also find out the regression equation using Excel.

5. Visit the website https://playground.tensorflow.org/ and execute classification or regression datasets. You can change the circled values and then click on play button to see the working of the model.



Each concept is the same as used in Teachable Machine.



6. Use the Linear Regression Calculator to find the 'Line of Best Fit'. Go to https://www.graphpad.com/quickcalcs/linear1/. Enter the values of x and y, then click on 'Calculate Now' to visualise the scatterplot.

Answers

Quiz

1. a 2. c 3. c 4. a 5. c 6. a 7. a 8. d 9. c 10. b

Exercise

A. 1. Machine learning 2. Dimensionality reduction 3. higher 4. Outlier
5. beforehand/prior 6. Anomaly detection 7. Distribution-based clustering
8. Clustering 9. Simple linear regression 10. supervised learning

B. 1. True 2. True 3. True 4. True 5. True 6. False 7. False
8. True 9. True 10. False

Scan the QR code or visit the following link to watch the video: Machine Learning in R—Classification, Regression and Clustering Problems -
https://www.youtube.com/watch?v=6za9_mh3uTE
After watching the video, answer the following questions:

1. Name the two applications of classification given in the video.

2. Also mention the two important points to keep in mind that the speaker mentioned about classification.

3. In Regression, the input variables are called ..... .

4. Clustering: grouping objects in clusters
   - ..... within clusters
   - ..... between clusters

5. Different clusters can reveal ..... and ..... information about the object.

UNIT-7

LEVERAGING LINGUISTICS AND COMPUTER SCIENCE

Learning Outcomes

- Challenges of the Human Language
- Natural Language Processing (NLP)
- Classification Problem
- Phases of NLP - Converting Speech to Text & analysing its intent
- IBM Project Debater
- Emotion Detection and Sentiment Analysis
- Chatbots
- Applications of NLP

Machines and humans are dynamically related. The existence of machines is not possible without human innovation and efforts and the purpose of machines fail if they don't fulfill human needs. Machines are designed to perform specific tasks to ease the work of humans, but they lack the understanding and adaptability of humans, this can lead to difficulties in communication, understanding and comprehension.

Human-machine interaction relies on words that may not always be intuitive or user-friendly, leading to frustration or inefficiency on the part of humans. Therefore, an interface that effectively bridge the gap between human intuition and machine functionality is crucial for enhancing interaction and usability.

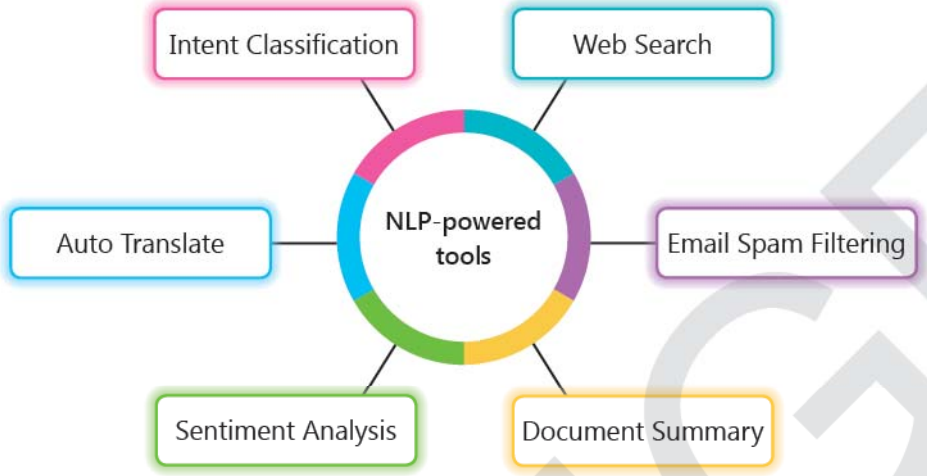
Challenges of the Human Language

Linguistics is a study of how to strategically apply language principles and practices to achieve specific aims or purposes. Linguistics is widely applied in marketing, advertising, communication, education, natural language processing, and other fields. Understanding how language works, (including structure, semantics, pragmatics, and sociolinguistics) enables individuals and organisations to modify messages, improve communication, and impact behaviour.

Understanding human language is a difficult task, especially for those who have grown up around it. Human language is extremely complex, with strange expressions that appear to contradict one another, vocabulary that take cultural knowledge to understand, and grammatical patterns that may turn simple statements into complex expressions or **tongue twisters**. Machines require systems known as Natural Language Processing (NLP) to understand human language.

NLP is a subfield of Artificial Intelligence (AI) that enables computers to understand, generate, and enhance human speech. NLP has the capacity to query data using natural language text or voice. Most of us have used NLP. For example, NLP is at the heart of the technology, that powers virtual assistants like as Oracle Digital Assistant (ODA), Siri, Cortana, and Alexa. NLP may be used on both written text and speech data.

Some examples of NLP-powered tools are as follows:



Brainy Fact

The Georgetown-IBM experiment, conducted on January 7, 1954, was a landmark example of machine translation. Georgetown University and IBM collaborated to develop the project, which involved complete automatic translation of more than sixty Russian sentences into English.

IBM Project Debater

Can an AI system maintain an ongoing conversation and respond to what is said with each exchange? Moreover, can an AI defeat a human expert in an argument over a complex topic? Consider a well-known experiment to find out how AI works with spoken and written language.

In 2012, IBM began developing **Project Debater** with the goal of creating a machine capable of more than just winning debates with humans. In other words, IBM Project Debater would listen to a series of competing arguments made by humans and respond to them intelligently, in addition to answering questions in human language. IBM's goal was to create a system that could assist individuals in making evidence-based, bias-free decisions on complex problems, where the answers were not evident.

A winning Project Debate could be prepared and presented by following the given steps:

1. **Understanding:** AI has huge data at its disposal, several billion passages from articles, magazines, newspapers, books, and journals. AI machine, on understanding the topic, create a vocabulary corpus and collate the data from its whole pool of sources.
2. **Reasoning:** After the analysis, a short speech is made from different text collected from varied sources with every minute detail presenting a compelling argument, in logical order, using good vocabulary.

3. **Learning:** Learn and understand the meaning of facts related to the topic. Gather and arrange facts supporting your argument and formulate your arrangement each time when new evidence arrives. This will help you find updated or completely new information that can score points against your opponent's position.
4. **Interacting:** Listening to your opponent's arguments and opinions, then present a convincing altercation that further proves your case.

Understanding

Cognitive systems understand like humans do.

Reasoning

They reason underlying ideas and concepts. They debate, infer and extract concepts.

Learning

They never stop learning. They develop "expertise" with every interaction and outcome.

Interacting

They interact with humans.

Ai Task

Communication

Find five interesting facts about "IBM Project Debater" and write in the space provided below. **[CBSE Handbook]**

Natural Language Processing (NLP)

Computers excel in working with structured data, in which everything is properly organised and labelled. Unfortunately for machines, human language is not structured. You've spent most of your life communicating through human language. Your brain accomplishes this with some of the most complex neural circuitry on this planet. However, creating machines that can understand human language, is extremely challenging.

NLP involves machines segmenting sentences and extracting meaning from "tokens" of human language. Human language is difficult and unstructured. Despite being loosely kept together by grammar constraints; our language presents information in a variety of puzzling ways. Unstructured information, as opposed to structured information, which may be organised in tables or matrices with clearly labelled rows and columns, is disorganised and difficult to comprehend. Consider the statements made by famous personalities:

- **Groucho Marx:** "One morning I shot an elephant in my pajamas. How he got into my pajamas I'll never know."
- **Steven Wright:** "I poured spot remover on my dog. Now he's gone."
- **George Carlin:** "I went to a bookstore and asked the saleswoman, 'Where's the self-help section?' She said if she told me, it would defeat the purpose."

To deal with the "messiness" of unstructured data, computers start with one sentence at a time. The process is known as **sentence segmentation** that is sentence-based text analysis. In NLP analysis, the text data is either analysed using meaningful words called **tokens**, or analysed using sentences. Computers divide the input into small piece of information (tokens), which may be classified separately. NLP can deal with text tokens after they have been sorted into a structured format or manner based on their meaning.

Entities, Relationships and Concepts

In natural language processing (NLP) and Artificial Intelligence, an **entity** is a significant piece of information or object in a text. An entity can be real-world objects such as names, locations, organisations, or dates. They can also be defined as a noun representing a person, place, a thing such as product name, technical phrase, or domain-specific concepts.

A **relationship** can be defined as a collection of two or more entities that share a strong bond with one another.

The following activities will show you how Groucho Marx's joke can be tokenised into useful categories called **entities** and **relationships**.

Ai Task

Communication

Now keeping this in mind, identify the entities in the sentence [CBSE Handbook]

"I shot an elephant in my pajamas."

Answer: I, elephant, pajamas

These relationships help explain the structure and humorous intent of the sentence.

Ai Task

Communication

Recall Groucho Marx's quote: [CBSE Handbook]

"One morning, I shot an elephant in my pajamas. How he got into my pajamas I'll never know."

Now keeping this in mind, identify the relationship between the entities in Groucho Marx's statement :

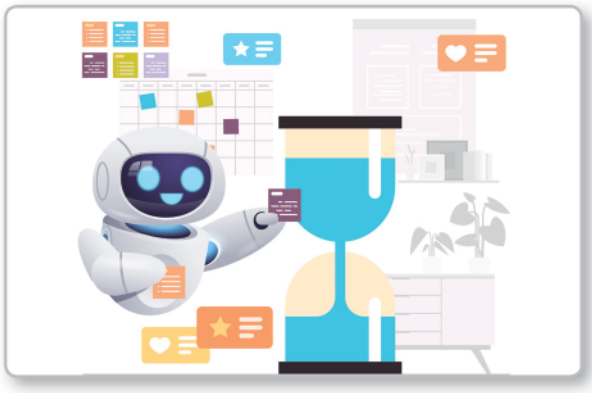
I + elephant, I + pajamas, in + pajamas, elephant + pajamas, I + shot

Answer: **Relationship between two entities** **Not a relationship between two entities**

I + elephant, I + pajamas

elephant + pajamas, I + shot **in + pajamas**

After an AI has identified entities and relationships within text or speech, it can start organising the information as part of its process to comprehend it. Interestingly, your brain performs a similar function, which may have assisted you in recognising entities and relationships in the above AI Task. Consider the following two sentences: "Siddharth ate the sweet dish. He always eats first." Notice the link between the two sentences: the word 'he' is linked to the term 'Siddharth'. The system identifies this association using understanding and linking different parts of the text known as **coreference resolution**. Coreference resolution is a crucial task in NLP where the system determines when different words refer to the same entity.



By recognising that 'he' refers to 'Siddharth', the AI can maintain coherence and context across sentences, much like how humans understand and relate different parts of a conversation or text.

A **concept** is an idea suggested or implied in a sentence but not actually stated. This is more challenging because it requires connecting underlying ideas rather than just the specific words used.

Ai Task

Communication

Recall, once more, Groucho Marx's statement: [CBSE Handbook]

"One morning, I shot an elephant in my pajamas. How he got into my pajamas I'll never know."

Based on this statement, select the words that are concepts implied, but not stated.

Safari	☐	Rifle	☐
Photographed	☐	Pajamas	☐

As AI continues to evolve, its ability to grasp and utilise implied concepts will become increasingly sophisticated, mirroring human cognitive processes and enhancing our interaction with technology.

Ai Task

Communication

Recall, once more, Steven Wright's statement: **"I poured spot remover on my dog. Now he's gone."**

1. Identify the entities in the statement.

   _____

2. State different relationship between the entities in the above statement.

   _____

3. Based on this statement, select the words show concepts implied, but not stated.

vanish	☐	cleansing agent	☐
dropping	☐	gone	☐

Ai Reboot

1. List 2 examples of NLP powered tools.

   _____

2. What was IBM's goal behind creating the Project Debater?

   _____

3. List the coreference resolution in "Adit is reading books. He loves to read."

   _____

4. Identify the entity, relationship and implied concept in the given statement: "Locked in a vault for fifty years, the owner decided to sell the jewels."

   _____

Emotion Detection and Sentiment Analysis

Sentiment analysis and emotion detection are two natural language processing (NLP) techniques that use human language to categorise people's thoughts, attitudes, and feelings. Sentiment analysis is a text classification technique that determines if subjective information is favourable, negative, or neutral. Emotion detection employs machine learning to examine complicated emotions such as fear, anger, sadness, love, and frustration.

At first look, sentiment analysis and emotion detection may appear to be the same concept particularly for individuals without a scientific background. However, they are not synonyms, so what is the difference between sentiment analysis and emotion detection?

Factors	Emotion Detection	Sentiment Analysis
Definition	Identifies various human emotion types.	Measures the intensity of an emotion.
Examples	Seeks to identify the emotions expressed in texts, such as happiness, rage, and grief.	Sentiment analysis seeks to categorise data as positive, negative or neutral.
Applications	Assessing user ratings, survey comments.	Reading social media content customer service chats etc.
AI Training	Can be taught to classify emotions.	Uses a sliding scale between positive and negative e.g. strongly disagree, disagree, neutral, agree, or strongly agree.
Purpose	Identifying emotional tokens to understand context.	Evaluating the overall tone or sentiment of text.

Sentiment analysis focuses on the overall sentiment expressed, whereas emotion AI seeks to distinguish and categorise distinct emotions. Thus, they work together to provide insights into the subjective components of human communication.

Classification Problem

In Natural Language Processing (NLP), classification problems include assigning a label or category to a piece of text, sentence, or phrase based on its content. For example, Emails can be classified as spam or non-spam, tweets as positive or negative, and articles as relevant or irrelevant to a specific topic.

Human language often includes terms that are vague or have multiple meanings. This leads to a classification problem. Understanding these terms can be challenging, especially for AI systems that rely on context to accurately classify given elements within sentences. Let us understand this with the help of an example. Consider the following statements:

- He's a really cool person. (Statement 1)
- The weather today is cool. (Statement 2)

In the above statements the phrases "cool person" and "cool weather" show the ambiguity of language, which poses a classification problem. In statement 1, "cool person" can refer to someone who is admired for their style or demeanour. Whereas in statement 2, a "cool temperature" refers to a lower, bearable level of heat. The challenge arises because the word "cool" has been used in two different ways.

Just as with humans, it can be tricky for an AI system to correctly interpret these meanings without proper context. Let's consider some additional examples:

- You can bank on Ravi to finish the project on time. (bank here means rely on)
- I need to go to the bank to deposit money. (here bank means a financial institution)

- She likes to rock in her chair while reading. (move)
- The band rocked the concert last night! (perform)

- You can store files on a cloud. (online storage)
- A cloud covers the sky. (weather phenomenon)

It might only take you a moment to understand these different meanings, but an AI system could struggle with classifying these elements correctly without a comprehensive understanding of language nuances and context.

Dealing with Classification Problems

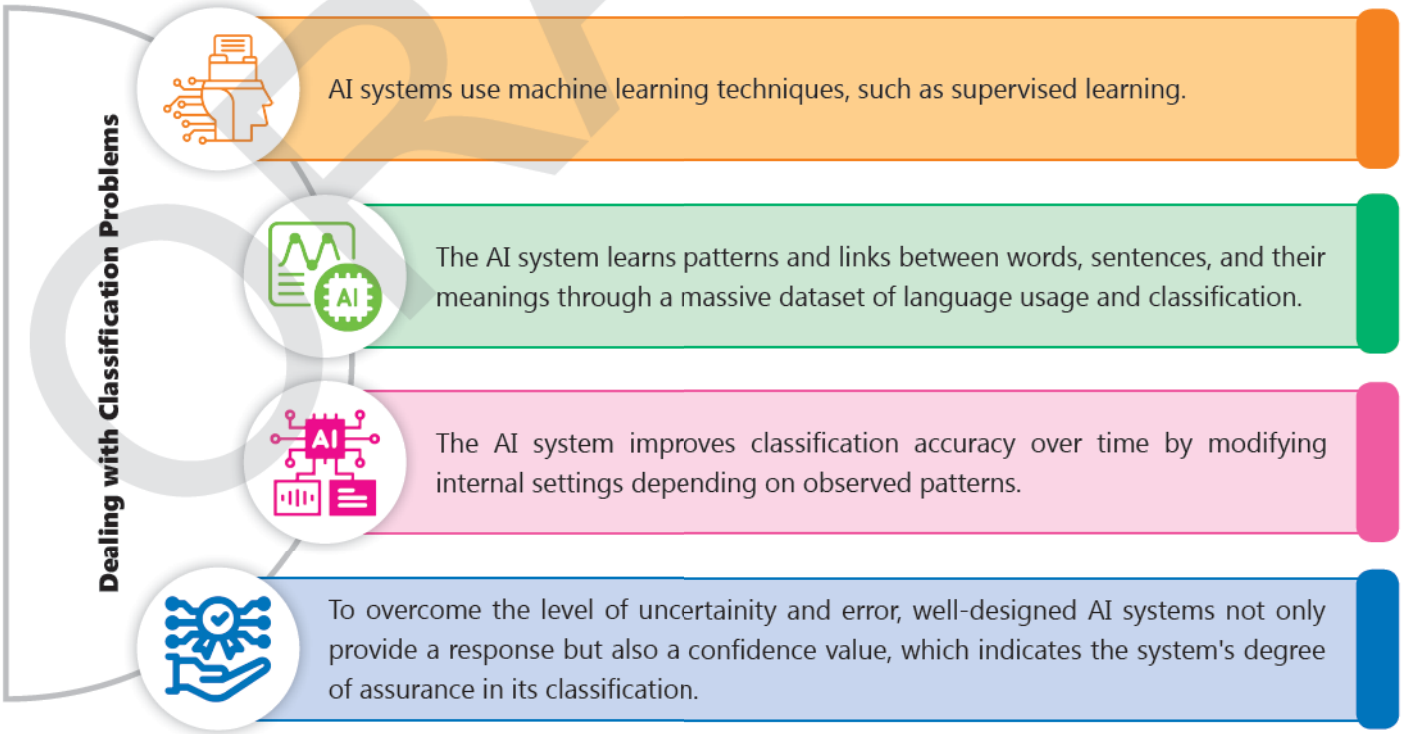
Classification can be more challenging for an AI system than simply identifying tokens because so much of classification depends on the context within the sentence.

Compare:

- I read a book about space. (outer space in universe)
- I need space to work. (personal space)

In both cases, the word "space" is used, but it has different meanings based on the context. An AI system must associate the word with the correct context: outer space or personal space.

How does an AI system deal with this problem? Here are some ways:



AI Task

Complete the analysis for each sentence, identifying the entities, relationships, and underlying concepts. This exercise will help you understand how AI and human brains process and relate information within a given context.

Sentence 1: Richa gave Aryan a book. He thanked her.

Entities: Richa, Aryan, Book, He (referring to Aryan), Her (referring to Richa)

Relationships: Richa → book → Aryan.

Aryan → thanked → Richa.

Concepts:

Coreference resolution: Recognising that "He" refers to "Aryan" and "her" refers to "Richa".

Gratitude: The idea that Aryan expressed thanks to Richa for the book.

Sentence 2: After winning the race, Surbhi was on cloud nine."

Entities: Surbhi, Race

Relationships: Surbhi → won → race.

Surbhi was extremely happy (implied by "on cloud nine").

Concepts:

Metaphor: "On cloud nine" implies extreme happiness rather than a literal location.

Chatbots

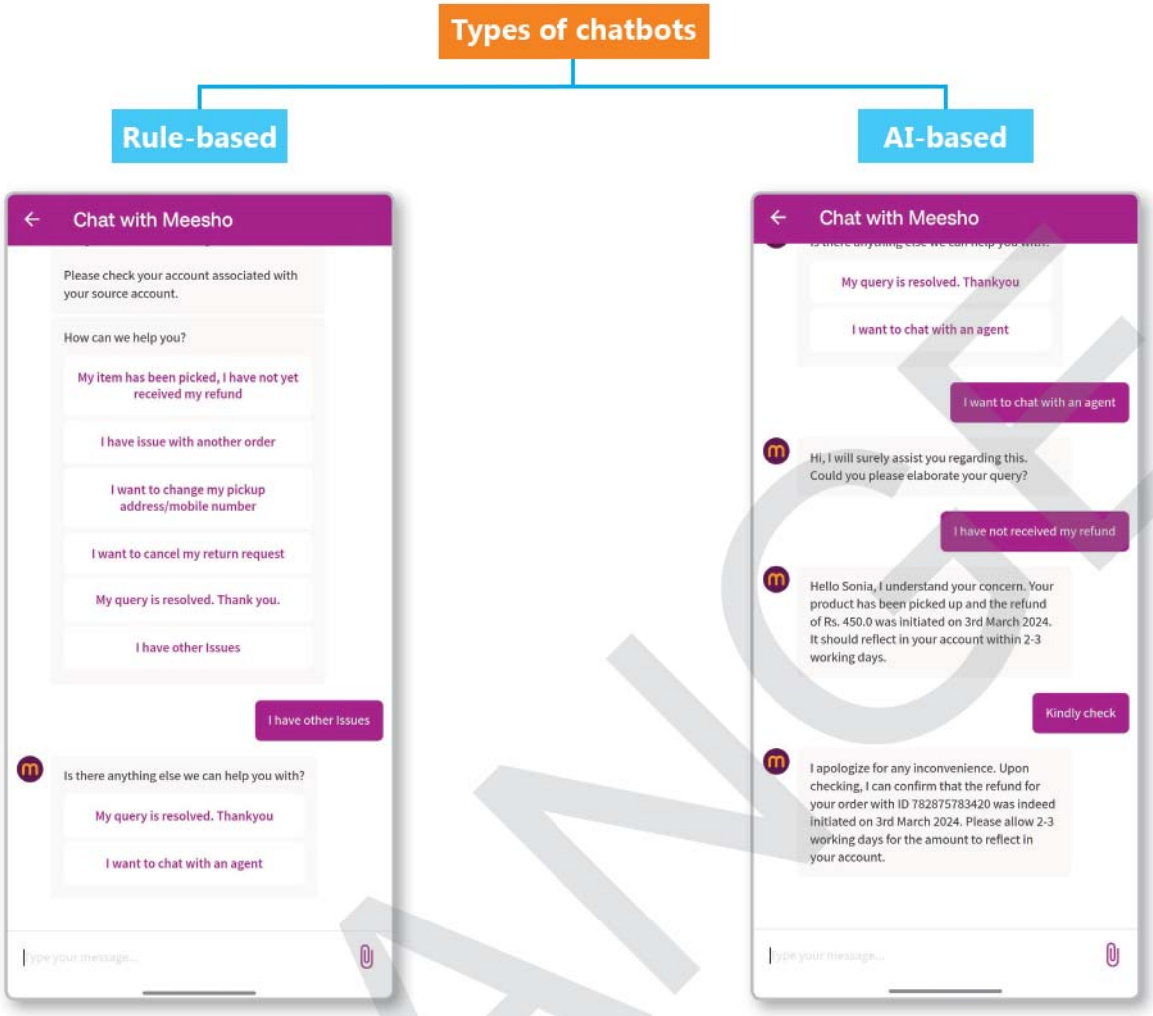
Imagine having a partner or a personal assistant who is always available for conversation, he never gets tired, and can provide instant answers to your questions. That's what a chatbot is! From ordering your favourite pizza to helping you complete your homework, chatbots are revolutionising how we interact with technology.

Chatbots are software applications or computer programs that simulate conversation with human user, usually via text or voice interactions. They use Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML) techniques to interpret user inquiries and respond appropriately. Chatbots can be connected to a variety of platforms, including websites, messaging apps, and voice assistants, and they can be used for customer service, information retrieval, task automation, and entertainment.

Types of Chatbots

Chatbots are used in a variety of industries from retail to healthcare, they provide constant availability for online engagement. While they are not always capable of providing solutions but they do listen and reply to repetitive requests, saving organisations from having dedicated human resources for such activities. *Chatbots work well with small amounts of data*, meeting specific demands such as movie queries for a cinema chain or broader inquiries for AI systems monitoring social media.

There are two types of chatbots:



Let us understand the working of each.

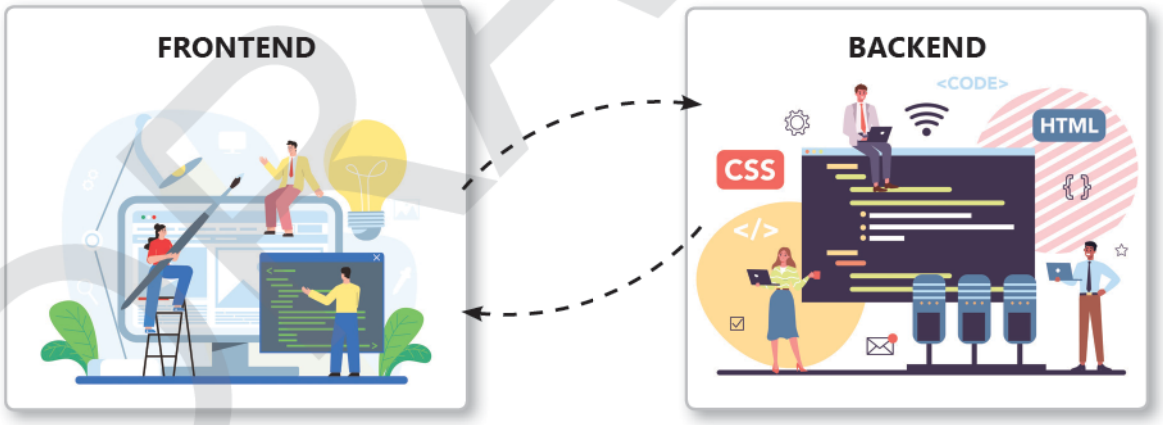
Factors	Rule-based Chatbot	AI-based Chatbot
Description	Work with established rules and decision trees. Respond to user input using pre-programmed rules.	Use Natural Language Processing (NLP) and Machine Learning methods. Also known as chat agents or virtual assistants.
Advantages	Simple to develop and maintain. Respond consistently and accurately to particular inquiries.	24-hour access for immediate and consistent assistance.
		Provide personalised interactions depending on users' preferences and history.
		Increase productivity and savings by automating tasks and lowering service costs.
Disadvantages	Struggle with understanding complex language.	Significant development expenditures and resource requirements.

Factors	Rule-based Chatbot	AI-based Chatbot
Disadvantages	Unable to adjust to conditions beyond predetermined rules.	Prone to biases in training data and a lack of transparency in decision-making.
		Ethical considerations for privacy, manipulation, and responsible use.
Applications	Customer service tasks include answering common questions and providing order status updates.	Entertainment and Gaming: Provide users with interactive stories and personalised gaming experiences.
	Guiding people through certain procedures.	Finance and Banking: Respond to questions regarding accounts, transactions, and financial products, as well as handle simple requests.

Structure of a Chatbot

A chatbot has both a "**frontend**" and a "**backend**".

A chatbot's **frontend** is the messaging channel through which users interact, and it has an easy-to-use interface. However, one drawback of the frontend is that it may lack contextual understanding, which means it may struggle to grasp the whole meaning or context of user communications. The **backend** of a chatbot is where the hard work happens. The backend handles application logic and has enough memory to remember previous sections of the discussion as it progresses.



Chatbots understand a query by breaking it down into simpler components and connecting them to objects in their memory. A chatbot's purpose is to recognise entities and intents and then use what it discovers to initiate a communication.

Intent

An **intent** is a purpose that explains why a user is contacting the chatbot. Consider it as a verb, or a type of activity. When communicating with a chatbot, users may have a variety of intentions, such as registering a complaint, requesting directions, or conversing with a sales representative. Organisations frequently have various intents for their chatbots to address.

For instance, assume you're developing a chatbot for a hotel chain. A common user's intent would be to ask about room availability. To ensure the chatbot can effectively handle this intent, you might ask a front desk agent who frequently answers these questions in person or over the phone. Your objective would be to compile a comprehensive list of the various ways guests might inquire about room availability. The table below includes different examples of potential user inputs related to this intent.

Intent	Possible user inputs
Check Room availability	Do you have any rooms available?
	Can I book a room for this weekend?
	Do you have any available rooms for next week?
	Is there a room available for two people?
	Are there extra charges for kids?
	Kids of what age group can stay in the same room?
	I want to book a room in your hotel. Is it available this weekend?

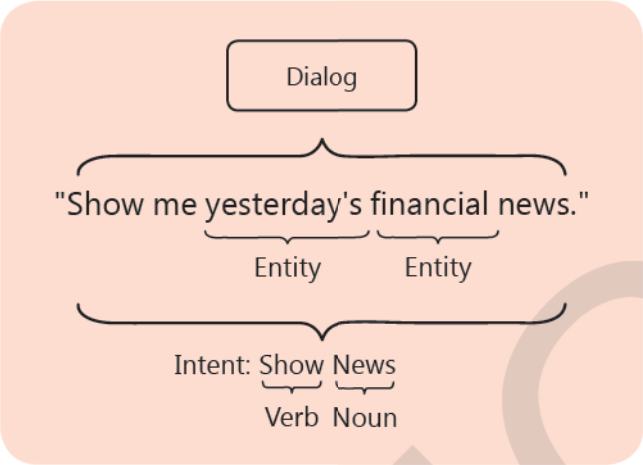
Entity

An entity is a noun that represents a person, place, or thing. Once you've compiled a list of the intents you want your chatbot to satisfy, you proceed further. For instance, if a user asks, "Do you have any rooms available in Delhi?", then checking room availability is the intent and Delhi is the entity. A chatbot requires a complete set of entities in order to be useful. The table below contains instances of entities that correspond to the intent and probable user inputs in the previous hotel example:

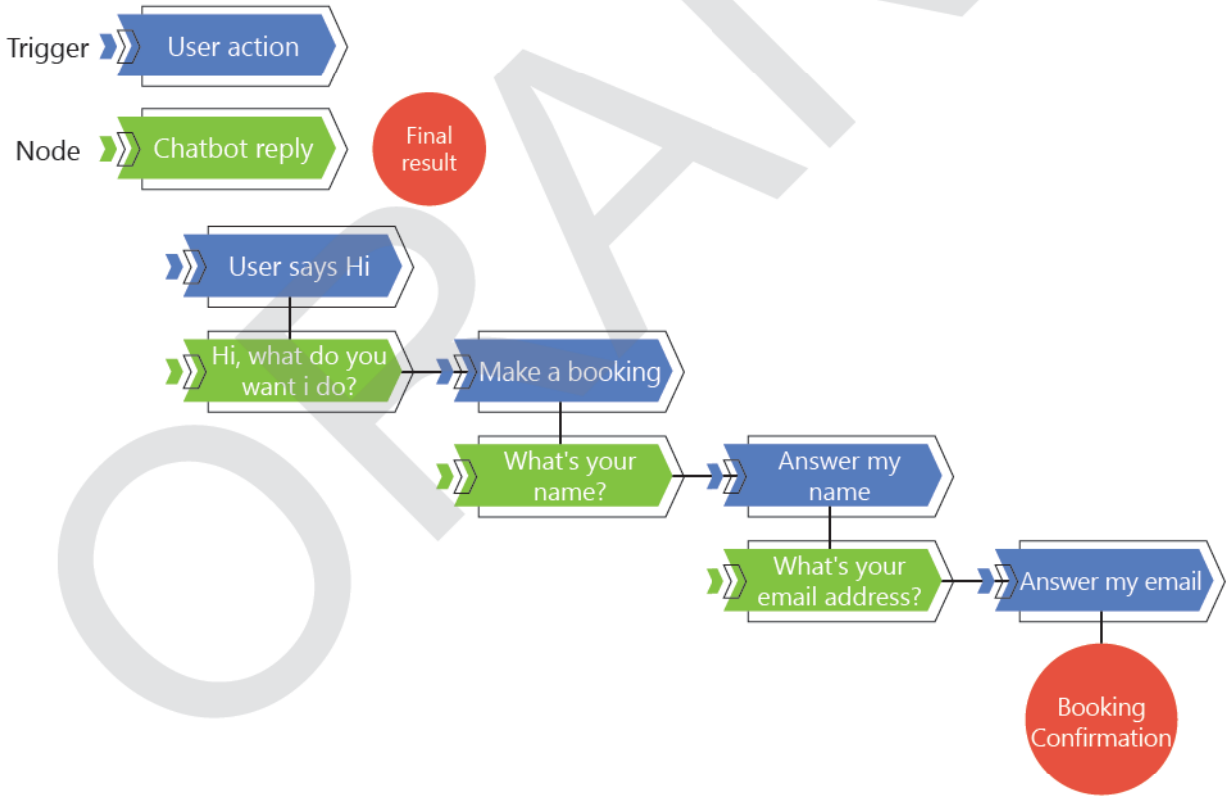
Intent	Possible user inputs	Entities
Check Room Availability	Do you have any rooms available in Delhi?	Location, Delhi
	Is there any room free for tonight?	Date
	Is there a room available for two people?	Number of guests
Book a Room	Can I book a room in Mumbai for next week?	Location and Date
	Can I reserve a room for tomorrow?	Date
	I want to book a family room for this weekend.	Room Type and Date

Dialog

A dialog is a flowchart – An IF / THEN tree structure that shows how a computer will react to user intents. When a human asks a query, the machine responds with a dialog. Even if a human uses run-on sentences, poor grammar, chat messaging expressions, and so on, then artificial intelligence allows the NLP to understand well enough to provide a response.



The dialog encompasses every possible word or phrase a user might use. The chatbot's matched response, and the various potential follow-up replies the user. Representing this complexity in a standard flowchart is challenging so chatbot software simplifies each conversation moment into a node. Each node includes a chatbot statement and an extensive as well as expandable list of potential user replies.



Designing the flowchart for a chatbot's responses is a detailed process. Each potential user input after the chatbot's initial greeting must be paired with an appropriate reply. For example, when users inquire about a hotel's room availability, any related questions would prompt the same response. This mapping of numerous possible questions to a few specific responses continues throughout the entire conversation.

Brainy Fact

According to Dashbot.io, a bot analytics service, the two most popular words to start a discussion with a chatbot are "hi" and "hello." A question mark(?), "hey," "help," "yes," and a thumbs up icon were among the most common messages. By the year 2029, it is expected that chatbots would acquire human-level language abilities.

AI Reboot

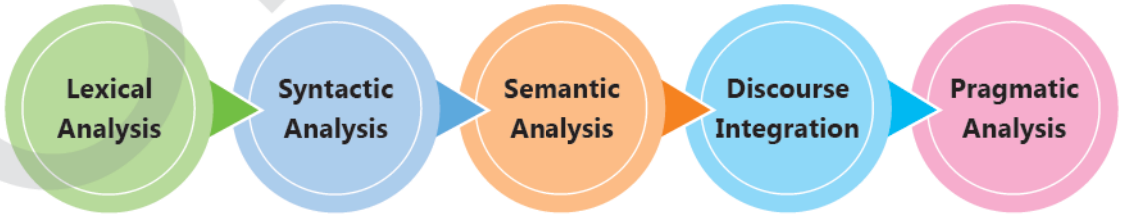
1. Fill in the blanks:
   a. AI that can be taught to identify emotions is called _____.
   b. The AI system learns _____ and links between words, sentences, and their meanings through a massive dataset of language usage and classification.
   c. _____ are software applications or computer programs that simulate conversation with human user, usually via text or voice interactions.
   d. A _____ chatbot struggles with understanding complex language.
   e. An intent is a _____ that explains why a user is contacting the chatbot.
   f. A dialog is a flowchart – An IF / THEN _____ structure that shows how a computer will react to user intents.
2. Write two possible queries which you will input in a chatbot when you want to inquire about booking tickets to a movie.

_____

_____

_____

Phases of NLP - Converting Speech to Text & analysing its intent

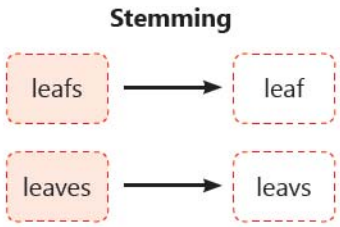


Natural Language Processing (NLP) involves a series of five phases that enable machines to analyse, categorise, and understand both spoken and written language. Companies use tools and algorithms that are in line with these NLP stages to derive insights from huge volumes of data and support well-informed decision-making. For NLP to be used effectively in text and voice applications, it is essential that these levels of NLP be understood. Let's explore each stage in greater detail.

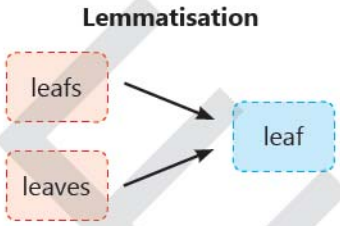
Lexical Analysis

This step involves studying and identifying a language's word structure. The text is divided into words, phrases, and paragraphs. Words are frequently reduced to their most basic forms using lexical normalisation techniques like lemmatisation and stemming.

- **Stemming** reduces words to their root form, such as removing suffixes like "ing", "ly", "es", and "s".



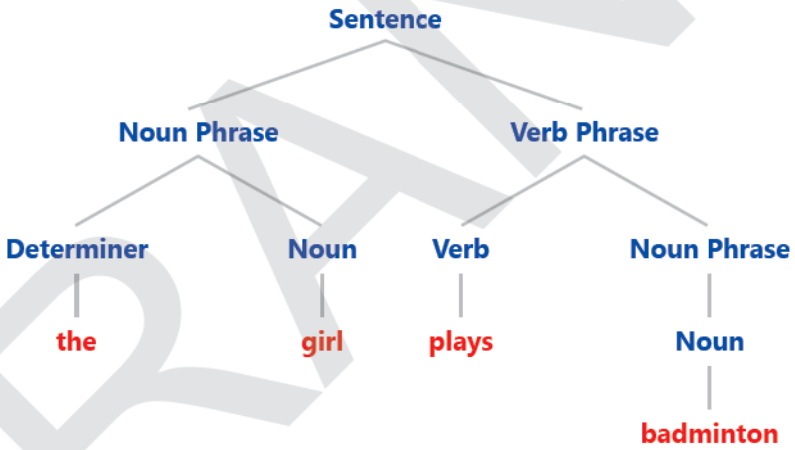
- **Lemmatisation** reduces words to their dictionary form, considering factors like **parts of speech** (POS) to determine their meaning in context.



Syntactic Analysis

Syntactic analysis is the second stage of NLP. Syntactic analysis, often known as **parsing**, is the act of examining grammar, word arrangement, identifying word relationships and determining if the sentence make sense. The approach involves examining all the words and phrases in a sentence as well as the structures that connect them.

A **syntax tree** is created as part of the procedure to visually represent semantic links. This method ensures that the structure, sequence, and grammar of sentences to make sense. Syntactic analysis also includes tagging words and sentences with POS (Parts of Speech) tags.



For Example: Badminton plays the girl: makes no sense, hence it is rejected in this stage. To analyse the words properly in a sentence for grammar we use **Syntactical parsing**. Dependency Grammar and Part of Speech (POS) tags are significant syntactic elements. The image given shows the correct syntactical parsing of words done during this stage.

Semantic Analysis

Semantic analysis is the third stage of NLP, which aims at determining the appropriate meaning behind a statement in a clear and concise manner. A semantic system integrates entities, concepts, relations, and predicates to add more context to language, allowing machines to understand text data more accurately.

Semantic analysis extracts meaning from language and serves as the foundation for a semantic system that aids robots in interpreting meaning. This involves mapping the syntactic structure and ensuring that the presented relationships between entities, words, phrases, and sentences in the text are logical.

In Natural Language, the meaning of a word may vary as per its usage in sentences and the context of the text. Semantic analysis involves interpreting the meaning of a word based upon the context in which it is used.

For example, the word 'Right' may mean 'direction' or 'correct.'

Similarly the word 'Band' may mean 'wrist band' or 'a group of people playing music ' – hence, the accurate meaning of the word is highly dependent upon its context.

Discourse Integration

Discourse integration is the fourth phase of NLP, which essentially implies contextualisation. Discourse integration is the process of analysing and identifying the bigger context for a smaller section of natural language structure (such as a phrase, word, entity or sentence).

During this phase, it is critical to ensure that each phrase, word, entity or sentence is mentioned in the proper context. This analysis considers sentence structure and semantics as well as sentence combination and overall text meaning.

For example: Radha wants it.

We can observe from the above sentence that the "it" keyword makes no sense. Hence, this statement is discarded. In reality, it applies to anything we don't know. Here "it" word depends on the prior sentence, which is not provided. So, if we know what "it" is, we can simply find the reference the girl is referring to.

Here its not clear what does Radha wants. It may be the laptop or cake or headphone or tablet. Its not specified what is 'it' referring to.

Pragmatic Analysis

Pragmatic analysis is the fifth and final stage of NLP. It is the process of abstracting or extracting meaning from the usage of language and interpreting a text using the knowledge obtained from previous NLP procedures. During this stage, understanding the meaning of words and the context in which they are used is completed. This allows for conversational functions between machines and humans (for example, chatbots). It translates the given text using the knowledge gathered in the preceding stages.

For Example:

Part 1	Part 2
The crew could see the whale.	which has surfaced only 50m behind them.

The second part of the sentence "which has surfaced only 50m behind them." is not making any sense unless and until the first part is known.

Ai Task

Collaboration and Teamwork

Read the following sentences and classify the meaning of the underlined words based on the context. You can discuss the meanings in groups of 2. This activity will help you understand how context plays a crucial role in classification problems, both for humans and AI systems.

1. She booked a room for the night.
2. He loves reading a good book.
3. They are conducting a large-scale survey.
4. He admired the survey of the landscape.

Video Session

Digital Literacy

Watch the following video to understand about the phases of NLP

Five Phases of NLP (Natural Language Processing) (with Tools and Applications)

https://www.youtube.com/watch?v=8SSAliFwCy8

1. Which phase interested you the most?

2. Which phase according to you is the most complex?

Brainy Fact

Remember 'Jarvis' in the Iron Man movies? The company, Hypersonix, has an AI-powered intelligent assistant, 'Jarvix'. Jarvix provides decision-makers with both prescriptive and predictive information so that they can make faster, better data-driven decisions. The company offers an NLP (Natural Language Processing) based research experience where you can ask questions orally or textually in simple English. Think of it as a 'Google for Enterprise' user experience designed to provide business users with a conversational experience.

Applications of NLP

Businesses can utilise Natural Language Processing tools to analyse data, discover insights, automate time-consuming tasks, and gain a competitive advantage.

Sentiment Analysis

Natural language processing (NLP) may analyse consumer comments, social media postings, product reviews, and other text data to assess positive, negative, or neutral sentiment towards a brand, product, or service. This data helps businesses understand their customers' impressions and sentiments.

Negative

Totally dissatisfied with the service. Worst customer care ever.

Neutral

Good Job but I will expect a lot more in future.

Positive

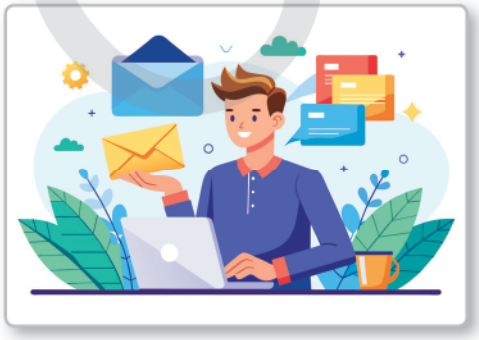
Brilliant effort guys! Loved Your Work.

Sentiment analysis

Generally, 80% of data is unstructured. That data needs to be analysed and used in a structured manner whether it is in the form of emails, texts, documents, articles, or anything else.

Voice Assistants

The use of voice assistants is growing in popularity! Almost everyone utilises a virtual assistant (like Siri, Alexa, or Google Assistant) for various tasks like making calls, setting alarms, scheduling meetings, making notes, and browsing the internet. How do they function? To comprehend and react to human speech, they employ a complex fusion of speech recognition, natural language processing, and natural language interpretation. The ultimate goal of voice assistants is to operate as an interface between people and the internet, offering a range of services using basic speech recognition.

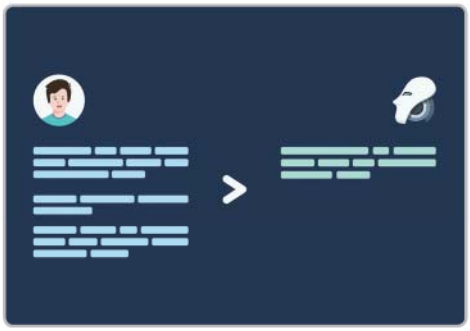


Email Filtering

We all use email on a regular basis. We get bombarded with communications concerning our studies, careers, and other subjects. We receive emails from a wide range of sources; some are promotional while others are work-related or from our ideal organisation. NLP is useful in this situation. It assigns emails based on whether they are considered "important" or "spam".

Document Analysis

Another use of Natural Language Processing is document analysis. Data is continuously being generated and needs to be properly kept, sorted, and searched in various fields including businesses, institutions, and schools. NLP has the potential to help with all of this. It saves us from the tiresome and time-consuming chore of looking through numerous files for a single person's information by not only searching a keyword but also categorising it according to the guidelines. Not only that, it also helps users make well-informed decisions on risk management and claims.



Automatic Summarisation

The scope of data processing has expanded with the growth of data. Processing data by hand takes a lot of time and is prone to mistakes. NLP offers a solution for this as well; it is able to extract the emotional meaning concealed in data while also summarising its significance. The summary procedure thus becomes more accurate and efficient.

Ai Reboot

1. Identify the NLP stage:
   a. This step involves studying and identifying a language's word structure.

   _____

   b. This involves mapping the syntactic structure and ensuring that the presented relationships between entities, words, phrases, and sentences in the text are logical.

   _____

   c. In this stage, grammar, word arrangement, identifying word relationships are examined to determine if they make sense.

   _____

   d. During this stage, understanding the meaning of words and the context in which they are used is completed.

   _____

2. Fill in the blanks:
   a. Generally, _____ % of data is unstructured.

   b. _____ helps users make well-informed decisions on risk management and claims.

   c. The ultimate goal of _____ is to operate as an interface between people and the internet, offering a range of services using basic speech recognition.

   d. _____ extracts the emotional meaning concealed in data along with its summarization keeping its significance.

Experiential Learning

Video Session

Check out https://sites.research.google/versebyverse/

This is an experimental AI-powered muse that helps you write poetry inspired by classic American poets.

Create your own poem on "Roses" or "A clear blue sky"

For Advanced Learners

Program 1: Create a simple Python code to display the POS (Parts of Speech) tags

```
# First, make sure to install nltk by executing following steps"
# open cmd prompt and type
# py -m pip install nltk
import nltk
from nltk.tokenize import word_tokenize
from nltk import pos_tag
# Download necessary NLTK data files
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
# Sample sentence
sentence = "The elephant is a majestic animal"
# Tokenize the sentence into words
words = word_tokenize(sentence)
# Perform part-of-speech tagging
pos_tags = pos_tag(words)
# Print the tokens with their respective parts of speech
for word, pos in pos_tags:
    print(f"Word: {word}, POS: {pos}")
```

Output:

```
Word: The, POS: DT
Word: elephant, POS: NN
Word: is, POS: VBZ
Word: a, POS: DT
Word: majestic, POS: JJ
Word: animal, POS: NN
```

Program 2: Create a simple chatbot using Python:

```
def get_response(user_input):
    if "hello" in user_input.lower():
        return "Hi there! How can I help you?"
    elif "how are you" in user_input.lower():
        return "I'm just a bot, but I'm here to help you!"
    elif "bye" in user_input.lower():
        return "Goodbye! Have a nice day!"
    elif "what is your name" in user_input.lower():
        return "I'm a simple chatbot created to assist you."
    elif "what can you do" in user_input.lower():
       return "I can respond to simple greetings and questions. Try asking me something!"
    else:
        return "I'm sorry, I don't understand."
def main():
    print("Welcome to the Simple Chatbot!")
    print("Type 'bye' to exit.")

    while True:
        user_input = input("You: ")
        if user_input.lower() == 'bye':
            print("Chatbot: Goodbye! Have a nice day!")
            break
        else:
            response = get_response(user_input)
            print("Chatbot:", response)

if __name__ == "__main__":
    main()
```

Output:

```
Welcome to the Simple Chatbot!
Type 'bye' to exit.
You: hello
Chatbot: Hi there! How can I help you?
You: how are you
Chatbot: I'm just a bot, but I'm here to help you!
You: what can you do
Chatbot: I can respond to simple greetings and questions. Try asking me something!
You: bye
Chatbot: Goodbye! Have a nice day!
```

Ai Task

Digital Literacy

Learn to create your first chatbot using botsify.com!

Watch the following videos:

- Botsify Tour In 90 Seconds - https://www.youtube.com/watch?v=Q0_xuMToKQM
- How To Create A Weather Bot With Botsify - https://www.youtube.com/watch?v=G9oxXIMxREw

At a Glance

- Machines rely on human efforts and serve human needs, but lack human understanding.
- Natural Language Processing (NLP) enables machines to understand and interact with human language.
- IBM Project Debater demonstrates AI's ability to engage in complex conversations and debates.
- NLP involves segmenting sentences and extracting meaning from unstructured human language.
- Emotion detection and sentiment analysis are distinct NLP techniques, offering insights into human communication.
- Classification problems in Natural Language Processing (NLP) involve assigning labels to text, sentences, or phrases based on their content.
- NLP is a subfield of Artificial Intelligence (AI) that enables computers to understand, generate, and enhance human speech.
- A winning Project Debate can be prepared and presented by following understanding, reasoning,learning, and interacting.
- An entity can be real-world objects such as names, locations, organisations, or dates.
- A relationship can be defined as a collection of two or more entities that share a strong bond with one another.
- Understanding and linking different parts of the text is known as coreference resolution.
- A concept is an idea suggested or implied in a sentence but not actually stated.
- Sentiment analysis and emotion detection are two Natural Language Processing (NLP) techniques that use human language to categorise people's thoughts, attitudes, and feelings.
- Sentiment analysis is a text classification task that determines if subjective information is favourable, negative, or neutral.
- Emotion detection employs machine learning to examine complicated emotions such as fear, anger, sadness, love, and frustration.
- Classification problems include assigning a label or category to a piece of text, sentence, or phrase based on its content.
- Chatbots are software applications or computer programs that simulate conversation with human user, usually via text or voice interactions.
- A chatbot's frontend is the messaging channel through which users interact, and it has an easy-to-use interface.
- The backend of a chatbot is where the hard work happens.
- An intent is a purpose that explains why a user is contacting the chatbot.
- An entity is a noun that represents a person, place, or thing.
- A dialog is a flowchart is an IF / THEN tree structure that shows how a computer will react to user intents.

- Natural Language Processing (NLP) involves a series of five phases that enable machines to analyse, categorise, and understand both spoken and written language.
- Lexical analysis involves studying and identifying a language's word structure.
- Syntactic analysis, often known as parsing, is the act of examining grammar, word arrangement, identifying word relationships and determining if they make sense.
- Semantic analysis aims at determining the appropriate meaning behind a statement in a clear and concise manner.
- Discourse integration is the process of analysing and identifying the bigger context for a smaller section of natural language structure.
- Pragmatic analysis is the process of abstracting or extracting meaning from the usage of language and interpreting a text using the knowledge obtained from previous NLP procedures.
- Businesses can use Natural Language Processing tools to analyse data, discover insights, automate time-consuming tasks, and gain a competitive advantage.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AiQuiz

A. Tick (✓) the correct option.

1. What is a primary reason for the dynamic relationship between machines and humans?
   a. Machines are self-sufficient
   b. Machines need human efforts for their creation
   c. Machines are designed to be adaptable without human intervention
   d. Machines do not require human input to perform tasks

2. In which fields is linguistics widely applied according to the text?
   a. Marketing, advertising, communications, and education
   b. Mathematics, physics, chemistry, and biology
   c. Engineering, medicine, law, and architecture
   d. Agriculture, astronomy, geology, and meteorology

3. What makes understanding human language difficult for machines?
   a. The simplicity of human language
   b. The uniform structure of human language across different cultures
   c. Strange expressions, cultural knowledge, and complex grammatical patterns
   d. The lack of vocabulary in human language

4. What should be done in the learning step while preparing a debate?
   a. Create a vocabulary corpus
   b. Present a convincing argument
   c. Gather and arrange facts supporting your argument
   d. Listen to your opponent's arguments

5. What role does cognitive system play in the debate preparation process?
   a. It only learns new information
   b. It helps to understand, reason, learn, and interact
   c. It focuses solely on understanding the topic
   d. It ignores opponent's arguments

6. According to Groucho Marx's statement : which of the following can be tokenized into useful categories?
   a. Adjectives and adverbs
   b. Entities and relationships
   c. Pronouns and prepositions
   d. Sentences and paragraphs

7. What is a key characteristic of a rule-based chatbot?
   a. Uses Natural Language Processing and Machine Learning
   b. Works with established rules and decision trees
   c. Provides personalised interactions based on user preferences
   d. Prone to biases in training data

8. Which technology does AI-based chatbots use to understand and respond to user input?
   a. Pre-programmed rules
   b. Natural Language Processing (NLP) and Machine Learning
   c. Decision trees
   d. Manual programming

9. What ethical considerations are associated with AI-based chatbots?
   a. Difficulty in understanding complex language
   b. Significant development expenditures
   c. Privacy, manipulation, and responsible use
   d. Consistent and accurate responses

10. Which of the following integrates entities, concepts, relations, and predicates to add more context to language, allowing machines to understand text data more accurately?
    a. Semantic System
    b. Pragmatic Analysis
    c. Discourse Integration
    d. Syntactic Analysis

11. Which phase of NLP involves studying and identifying a language's word structure and breaking down text into words, phrases, and paragraphs?
    a. Syntactic Analysis
    b. Lexical Analysis
    c. Semantic Analysis
    d. Pragmatic Analysis

12. What does discourse integration primarily focus on?
    a. Extracting meaning from word usage in context
    b. Analysing and identifying the bigger context for smaller sections of text
    c. Reducing words to their root forms
    d. Ensuring grammatical correctness of sentences

13. Which stage of NLP translates the given text using the knowledge gathered in the preceding stages?
    a. Sentiment Analysis
    b. Syntactic Analysis
    c. Lexical Analysis
    d. Pragmatic Analysis

14. How does document analysis contribute to NLP applications?

a. Summarising documents automatically ○

b. Sorting and searching data in businesses, institutions, and schools ○

c. Translating documents into different languages ○

d. Categorising emails ○

B. Fill in the blanks.

1. _____ are designed to perform specific tasks to ease the work of humans.
2. Cognitive systems understand, reason, learn, and _____.
3. Listening to your opponent's arguments and opinions, then present a convincing _____ that further proves your case.
4. _____ work with established rules and decision trees and respond to user input using pre-programmed rules.
5. Voice Assistants, such as Siri, Alexa, or Google Assistant, utilise a complex fusion of _____, natural language processing, and natural language interpretation to comprehend and react to human speech.
6. In _____, NLP assigns emails based on whether they are considered "important" or "spam".
7. _____ saves us from the time-consuming chore of looking through numerous files by searching a keyword and categorising it.
8. _____ aims at determining the appropriate meaning behind a statement in a clear and concise manner.
9. The semantic analysis involves mapping the syntactic structure and ensuring that the presented relationships between entities, words, phrases, and sentences in the text are _____.
10. _____ is a purpose that explains why a user is contacting the chatbot, which can be considered as a verb or a type of activity.

C. State whether the following statement is True or False.

1. An entity can be real-world objects such as names, location, organisations or dates. _____
2. The process of understanding and linking different parts of the text is known as coreference resolution. _____
3. A concept is an idea suggested or implies in a sentence but not actually stated. _____
4. Sentiment analysis and emotion detection are two natural language processing (NLP) techniques that use human language to categorise people's thoughts, attitudes, and feelings. _____
5. Chatbots are software applications or computer programs that simulate conversation with machine. _____
6. A chatbot's backend is the messaging channel through which users interact and it has an easy-to-use interface. _____
7. An intent is a purpose that explains why a user is contacting the chatbot. _____
8. The text is divided into words and paragraph in the step of discourse integration. _____
9. The voice assistants are losing their popularity. _____
10. Processing data by hand takes a lot of time and is prone to mistakes. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. When and why did IBM begin developing Project Debater?

Ans. In 2012, IBM began developing Project Debater, with the goal of creating a machine capable of more than just winning debates with humans. IBM's goal was to create a system that could assist individuals in making evidence-based, bias-free decisions on complex problems where the answers were not evident.

2. Define entity and relationship.

Ans. An entity can be real-world objects such as names, locations, organisations, or dates. They can also be defined as a noun representing a person, place, a thing such as product name, technical phrase, or domain-specific concepts. A relationship can be defined as a collection of two or more entities that share a strong bond with one another.

3. Differentiate between frontend and backend of a chatbot.

Ans. A chatbot's frontend is the messaging channel through which users interact and it has an easy-to-use interface. However, one drawback of the frontend is that it may lack contextual understanding, which means it may struggle to grasp the whole meaning or context of user communications. The backend of a chatbot is where the hard work happens. The backend handles application logic and has enough memory to remember previous sections of a discussion as it progresses.

4. "Computers excel in working with structured data, in which everything is properly organized and labelled. Unfortunately for machines, human language is not structured." What is so different about the human language.

Ans. Human language is extremely complex, with strange expressions that appear to contradict one another, vocabulary that take cultural knowledge to understand, and grammatical patterns that may turn simple statements into complex expressions or tongue twisters.

5. Explain 'concept' with the help of an example.

Ans. A **concept** is an idea suggested or implied in a sentence but not actually stated. This is more challenging because it requires connecting underlying ideas rather than just the specific words used. Example - George Carlin said: "I went to a bookstore and asked the saleswoman, 'Where's the self-help section?' She said if she told me, it would defeat the purpose."

B. Long answer type questions.

1. What are the steps for preparing and presenting a winning Project Debate?

Ans. A winning Project Debate can be prepared and presented following the given steps: -

- **Understanding:** AI has huge data at its disposal, several billion passages from articles, magazines, newspapers, books, and journals. AI machine on understanding the topic create a vocabulary corpus and collate the data from its whole pool of sources.
- **Reasoning:** After the analysis a short speech is made from different text collected from varied sources with every minute detail presenting a compelling argument, in logical order, using good vocabulary.
- **Learning:** Learn and understand the meaning of facts related to the topic. Gather and arrange facts supporting your argument and formulate your arrangement each time when new evidence arrives. This will help you find updated or completely new information that can score points against your opponent's position.
- **Interacting:** Listen to your opponent's arguments and opinions, then present a convincing altercation that further proves your case

2. Differentiate between Rule-based Chatbot and AI-based Chatbot.

Ans.

Factors	Rule-based Chatbot	AI-based Chatbot
Description	Work with established rules and decision trees. Respond to user input using pre-programmed rules.	Use Natural Language Processing (NLP) and Machine Learning methods. Also known as chat agents or virtual assistants.
Advantages	Simple to develop and maintain. Respond consistently and accurately to particular inquiries.	24-hour access for immediate and consistent assistance.
		Provide personalised interactions depending on users' preferences and history.
		Increase productivity and savings by automating tasks and lowering service costs.
Disadvantages	Struggle with understanding complex language.	Significant development expenditures and resource requirements.
	Unable to adjust to conditions beyond predetermined rules.	Prone to biases in training data and a lack of transparency in decision-making.
		Ethical considerations for privacy, manipulation, and responsible use.

3. Differentiate between Emotion Detection and Sentiment Analysis.

Ans.

Factors	Emotion Detection	Sentiment Analysis
Definition	Identifies various human emotion types.	Measures the intensity of an emotion.
Examples	Seeks to identify the emotions expressed in texts, such as happiness, rage, and grief.	Sentiment analysis seeks to categorise data as positive, negative or neutral.
Applications	Assessing user ratings, survey comments.	Reading social media content, customer service chats, etc.
AI Training	Can be taught to classify emotions.	Uses a sliding scale between positive and negative e.g. strongly disagree, disagree, neutral, agree, or strongly agree.
Purpose	Identifying emotional tokens to understand context.	Evaluating the overall tone or sentiment of text date.

4. Explain discourse integration with an example.

Ans. Discourse integration is the process of analysing and identifying the bigger context for a smaller section of natural language structure (such as a phrase, word, entity or sentence). During this phase, it is critical to ensure that each phrase, word, entity, or sentence is mentioned in the proper context. This analysis considers sentence structure and semantics as well as sentence combination and overall text meaning.

For example: Radha wants it.

We can observe from the above sentence that the "it" keyword makes no sense. Hence, this statement is discarded. In reality, it applies to anything we don't know. Here "it" word depends on the prior sentence, which is not provided.

5. "A syntax tree is created as part of the procedure to visually represent semantic links." Identify the phase of NLP processing? **[CBSE Handbook]**

Ans. The phase of NLP processing that involves the creation of a syntax tree to visually represent semantic links is the Syntactical Analysis phase.

In syntactical analysis, the aim is to check the grammar, word layouts, and word relationships in a given text. One of the key tasks in this phase is to create a syntax tree, also known as a parse tree, which represents the grammatical structure of the sentence and visually displays the relationships between words. This helps in understanding the syntactical constructs and semantic relationships within the text, thereby aiding in the overall comprehension and analysis of the language.

C. Competency-based/Application-based questions:

1. In "EduTech University", Professor Geeta envisioned an AI, capable of debating with human-like depth and knowledge. This idea sparked "Project Debator", motivated by a desire to enhance public discourse through unbiased, well-researched arguments.

   Professor Geeta wanted the AI to access vast information, articulate points clearly, and anticipate counter arguments. She also envisioned the AI model to learn from interactions, analyse emotional tones, and provide feedback to improve users' debating skills. With this vision, she and her team embarked on developing "Project Debator" to revolutionise how people think, argue, and understand the world. What features would you like to add on from your side, apart from features envisioned by Professor Geeta?

Ans. Professor Geeta's Project Debater aims to create an AI for unbiased debate by providing well-researched arguments, clear communication, and the ability to learn and adapt. To this, you propose adding features like:

- Creative argument approaches using metaphors and storytelling.
- Contextual understanding to tailor arguments for specific audiences.
- Real-time fact-checking for extra credibility.
- Fallacy detection to counter illogical arguments.

2. A ground-breaking AI project was underway at the esteemed "SoftCo Labs". Dr. Madhavi, the lead researcher, was deep in thought as she looked over her team's latest developments in Natural Language Processing (NLP). They had successfully integrated sentiment analysis into their system, but now faced a crucial decision: should they treat sentiment analysis and emotion detection as separate units?

   Gathering her team, Dr Madhavi posed a critical question. "Imagine our AI, Ava, navigating the complex world of human interaction. What reasons would justify us dividing sentiment analysis and emotion detection into two distinct units of NLP?"

   As her team pondered this, they knew they had to consider various factors to ensure Ava could understand and respond to human communication with the utmost accuracy and empathy. What reasons would they come up with to support Dr. Madhavi's proposition?

Ans. Sentiment analysis and emotion detection are two Natural Language Processing (NLP) techniques that use human language to categorise people's thoughts, attitudes, and feelings. Sentiment analysis is a text classification task that determines if subjective information is favourable, negative, or neutral. Emotion detection employs machine learning to examine complicated emotions such as fear, anger, sadness, love, and frustration.

At first look, Sentiment Analysis (SA) and Emotion Detection (ED) may appear to be the same concept particularly for individuals without a scientific background. However, they are not synonyms, a simple distinction would be that emotion detection tells the emotion where as sentiment analysis detects the intensity. Emotion detection understands the tokens for emotions where as sentiment analysis looks for the tone to identify the feelings

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. What field of study is crucial for understanding and improving human language communication?
   a. Mathematics ○ b. Chemistry ○
   c. Linguistics ○ d. Physics ○
2. Which of the following is an example of a virtual assistant powered by NLP?
   a. Google Search ○ b. Microsoft Word ○
   c. Siri ○ d. Photoshop ○
3. What is the first step in preparing a winning Project Debate using AI?
   a. Reasoning ○ b. Interacting ○
   c. Understanding ○ d. Learning ○
4. Which of the following is required by machines to understand human language?
   a. NLP ○ b. Neural network ○
   c. Chatbot ○ d. Data parsing ○
5. Which of the following is used by Natural Language Processing (NLP) to assess customers' impressions and sentiments?
   a. Sentiment Analysis ○ b. Syntactic Analysis ○
   c. Lexical Analysis ○ d. Pragmatic Analysis ○
6. What is a relationship in the context of NLP?
   a. A single entity in a text ○
   b. A random word in a sentence ○
   c. A collection of two or more entities that share a strong bond ○
   d. An individual noun ○
7. What is an advantage of rule-based chatbots?
   a. They can adapt to conditions beyond predetermined rules. ○
   b. They provide personalised interactions. ○
   c. They are simple to develop and maintain. ○
   d. They are prone to biases in training data. ○
8. What is a notable advantage of AI-based chatbots over rule-based chatbots?
   a. They are easier to develop and maintain. ○
   b. They struggle with understanding complex language. ○
   c. They provide personalised interactions based on user preferences and history. ○
   d. They respond consistently and accurately to particular inquiries. ○
9. In Natural Language Processing (NLP), which of the following include assigning a label or category to a piece of text, sentence, or phrase based on its content?
   a. Classification problems ○ b. Chatbot ○
   c. Sentiment analysis ○ d. Emotion detection ○

10. Which of the following techniques is used in Lexical Analysis to reduce words to their most basic forms?

a. Parsing ◯ b. Discourse Integration ◯

c. Lemmatisation and Stemming ◯ d. Part of Speech Tagging ◯

11. How do chatbots fundamentally alter the corporate landscape?

a. By providing automatic document translation services ◯

b. By being more efficient than people and connecting with a wider audience on messaging apps ◯

c. By filtering emails more effectively than human agents ◯

d. By summarising long documents into shorter versions ◯

12. Which of the following applications of NLP is mentioned as helping users make well-informed decisions on risk management and claims?

a. Sentiment Analysis ◯ b. Document Analysis ◯

c. Email Filtering ◯ d. Voice Assistants ◯

B. Fill in the blanks.

1. Machines and humans are _____ related. The existence of machines is not possible without human efforts and purpose of machines fail if they don't justify human needs.
2. Auto-translate is an example of _____ tools.
3. _____ and _____ are similar but one identifies emotion and other finds out the emotional intensity.
4. _____ are also known as chat agents or virtual assistants.
5. Syntactic analysis includes tagging words and sentences with _____.
6. _____ employs machine learning to examine complicated emotions such as fear, anger, sadness, love, and frustration.
7. In order to design a complete structure of a chatbot we need _____, _____ and _____.
8. To overcome the level of uncertainity and error, well-designed AI systems not only provide a response but also a confidence value, which indicates the system's degree of _____ in its classification.
9. A chatbot's _____ is the messaging channel via which users engage, and it has an easy-to-use interface.
10. The dialog encompasses every possible _____ a user might use.

C. State whether the following statement is True or False.

1. Human language is extremely simple, with strange expressions that appear to synchronise with every other thing. _____
2. Understanding is listening to your opponent's arguments and opinions, then present a convincing altercation that further proves your case. _____
3. Entity can also be defined as a noun representing a person, place, a thing such as product name, technical phrase, or domain-specific concepts. _____
4. In coreference resolution, system determines, when different words refer to the same entity. _____
5. NLP is a subfield of Artificial Intelligence (AI) that enables computers to understand, generate, and enhance human speech. _____
6. Chatbots that work well with small amounts of data are called AI-based chatbot. _____
7. An entity is a verb that represents an idea suggested or implied as well as stated. _____
8. Lexical analysis is also known as parsing. _____

9. In natural language, the meaning of a word may vary as per its usage in sentences and the context of the text. _____

10. Pragmatic analysis is the process of abstracting or extracting meaning from the usage of language and interpreting a text using the knowledge obtained from previous NLP procedures. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. Explain Coreference resolution with an example.
2. Define the following:

   (a) Intent (b) Dialog (c) Tokens
3. Define any 2 applications of NLP.
4. "Human language is full of terms that are vague or have double meanings". Explain this statement with an example.
5. List the five phases of NLP.

B. Long answer type questions.

1. Discuss the application of document analysis.
2. Elaborate all the stages of NLP.
3. Explain the structure of a Chatbot with a suitable diagram.
4. Explain the concept of dialog with the help of an example.
5. What strategies are adopted by an AI model to deal with the classification problem?

C. Competency-based/Application-based questions:

1. In the town of Hawri , a young AI enthusiast named Alisha was developing a smart email filter. One day, she encountered a puzzling challenge and turned to her mentor, Dr. Lina, for advice.

   Alisha asked, "What kind of classification problems does an AI system face?"

   Dr. Lina smiled and replied, "Imagine your email filter needs to classify incoming messages as 'spam' or 'not spam.' What examples can you think of, that highlight the challenges it might face in this classification problem?". Explain your response with an example.
2. As the head of a retail customer support team, your team currently handles customer complaints and inquiries manually. To improve efficiency, you are considering implementing a chatbot (either rule-based or AI-based). Describe the challenges you might face during the transition and the expected impact on your team's efficiency and customer satisfaction.

Ai In Life — Creativity and Innovativeness

Voice assistants are likely to become more common in the future. They are already employed in many aspects of our lives, including smart homes, connected cars, and mobile devices, and they are likely to become considerably more common in the future. Predict a few characteristics that future voice assistants will undoubtedly have.

Ai Deep Thinking

Collaboration and Teamwork

Brainstorm and list challenges and concerns you can think of related to voice assistants. Think about privacy, security, reliability, accessibility issues, biases, and potential misuse. Present your findings to your class.

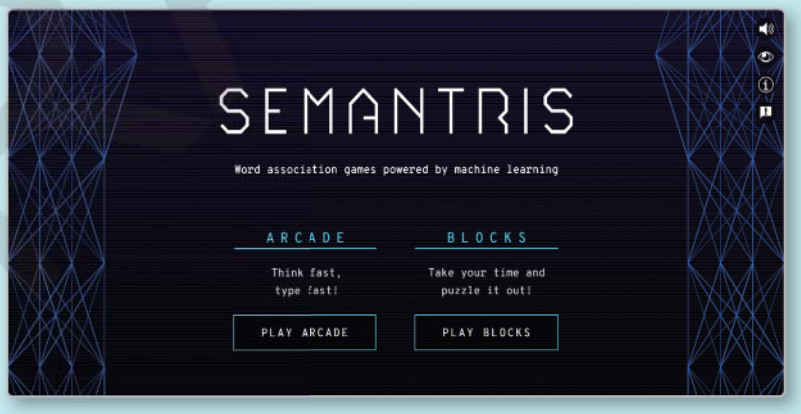
Ai Lab

Experiential Learning

1. Check out a free sentiment analysis tool! Go to https://text2data.com/Demo

Just type the text and let the AI categorise as positive, negative or neutral.

2. Play Semantris, a word association game developed by Google. Go to https://research.google.com/semantris/



Choose to play Arcade or Blocks! Enjoy!

1. Chat with an award winning chatbot, Kuki! Go to https://www.kuki.ai

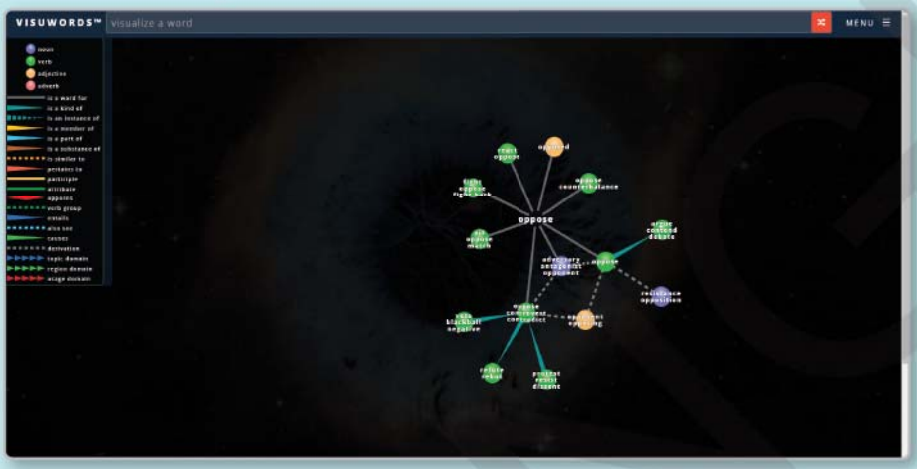
   Did you notice Kuki's sense of humour? What do you think her age would have been, had she been a human?

2. Create a pizza ordering chatbot using Google Dialogflow/botsify.com/botpress.com

Watch the following videos to learn to create a chatbot on Dialogflow

- Pizza bot with Dialogflow - https://www.youtube.com/watch?v=DifAemVDAis
- Creating ChatBot || Dialogflow || Just in 30 Minutes - https://www.youtube.com/watch?v=0SDfi3JvacE&t=27s

3. Ever heard of a visual dictionary? Check out https://visuwords.com. Type a word and see its different meanings!



Answers

Exercise (Section A)

A. 1. b 2. a 3. c 4. c 5. b 6. b 7. b 8. b 9. c 10. a 11. b 12. b 13. d 14. b

B. 1. Machines 2. interact 3. altercation 4. Rule based chatbot 5. speech recognition 6. Email filtering 7. Document analysis 8. Semantic analysis 9. logical 10. intent

C. 1. True 2. True 3. True 4. True 5. False 6. False 7. True 8. False 9. False 10. True

Scan the QR code to watch the following video - IBM Project Debater - We Should Subsidize Preschool

https://www.youtube.com/watch?v=-d4Uj9ViP9o&t=1462s

After watching the above video, answer the following questions:

1. Were the arguments given by the Project Debater as good as a human?

2. Was the Project Debater able to contextualise (put into perspective) the arguments in a well-structured manner?

3. Identify the metaphor used by Projector Debater. (Hint: suitcase) Do you think it was relevant in the debate's context?

4. What was the one thing that you liked about the Project Debater?

UNIT-8

AI ETHICS AND VALUES

Learning Outcomes

- Ethics In Artificial Intelligence
- The Five Pillars of AI Ethics
- Mitigating Bias in AI Systems
- AI and Ethical Concerns
- Moral Machine
- Why is AI Ethics Important?
- Bias, Bias Awareness, AI Bias and Sources of Bias
- Developing AI Policies
- Ethical Dilemma

Artificial Intelligence (AI) is pervasive in modern life. It powers navigation apps, helping you find the best or most eco-friendly routes. It enhances the capability of search engines, making it easier to find the information you need. It assists doctors in making accurate diagnoses and developing better treatment plans. It improves weather forecasting, so you can be prepared for major weather events. Combined with sensors and satellites, it collects data about the environment, helping scientists in understanding and predicting changes in our world. AI makes life easier and safer by helping people to make informed decisions, connecting them with the right information at the right time, and identifying patterns or efficiencies they might miss.

But AI can also cause harm. For instance, AI can be used to decide who gets a loan, who gets into college, who gets a job, how much employees are paid, or how long prison sentences should be. The harm from AI isn't always physical. It can be less obvious, like unfairness, discrimination, or exclusion. This harm can be subtle because people might not know when they are interacting with AI or how it is affecting decisions about them.

Artificial Intelligence is the product of human imagination. Prejudice or bias, whether we are willing to admit it or not, is inherent in human experience. To avoid prejudice, programmers can turn to the philosophical basis of asking questions out of curiosity, seeking to transcend our own perspectives, and learning to see the world outside the trap of our own human algorithms. Machines are everywhere and they learn from humans. Since all humans have certain ethics and biases based on the environment and cultures they are surrounded by, machines also learn the ethics and biases from humans.

As Artificial Intelligence (AI) evolves towards a stage where it can replicate human brain capabilities. The development of intelligent machines raises some ethical concerns. To make people aware about AI machines, some ethical guidelines have been developed. These guidelines ensure that AI machines are not harmful to humans or society.

Ai Task

Digital Literacy

Scan the QR code or go to the link to watch the video and answer the question given below:
AI is Monitoring You Right Now and Here's How It's Using Your Data

https://www.youtube.com/watch?v=KpybityrXfs

From where do computer forecasting systems collect data?

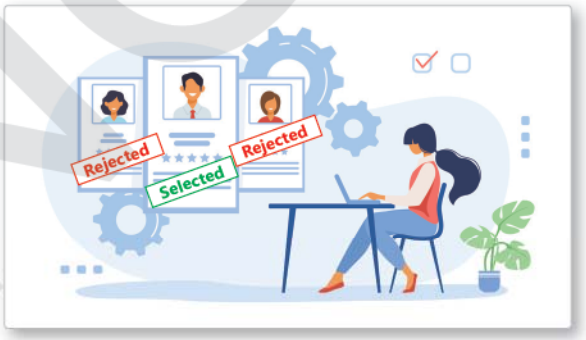
_____

Ethics In Artificial Intelligence

AI ethics is a set of moral principles and practices designed to support the development and responsible use of AI technology. As AI becomes an indispensable part of products and services, companies have begun to develop AI ethics guidelines. The aim is to guide stakeholders when faced with ethical dilemmas while using AI.

Case 1: The Biased Hiring Tool

Sakshi is excited about applying for a software engineering position at a top tech company. She has all the qualifications and experience needed for the role. Unknown to her, the company uses an AI tool to screen applicants. The AI, trained on a decade of resumes mostly from men, has a hidden bias against women. As a result, Sakshi's application is not rated properly, ranking lower than those of less qualified male candidates. Despite her skills and experience, she never gets a call for an interview. The company missed a talented candidate, and Sakshi is left wondering what went wrong.



Due to biased judgment of AI, a deserving candidate is left out. Think, is it fair to depend totally on an AI tool or, if we are depending on it, is it giving us fair results?

Case 2: The Autonomous Car's Dilemma

Ritesh owns an autonomous car manufactured by "NEWAR" cars that promises safety and convenience. One rainy night, the car encounters an unexpected obstacle on the road. It must make a split-second decision: swerve onto a busy sidewalk or stay on course and risk a collision with a dog. The AI has been programmed to minimise harm, but in this scenario, there is no perfect solution. Ritesh, a bystander, and the other driver all face potential danger due to the AI's decision-making limitations.

Who do you think is wrong in the given case? Is it Ritesh owner of the autonomous car, the NEWAR car manufacturing company, the technician who designed the car, or the bystanders for being on the road?

Ethical & Moral Reasoning

Video Session

Scan the QR code or visit the following link to watch the video: *The Ethical Robot*

https://www.youtube.com/watch?v=pajCoSTGvas

After watching the video answer, the following question:

What were the ethical concerns expressed in the video?

Ethical & Moral Reasoning

Video Session

Scan the QR code or visit the following link to watch the video:
The Three Big Ethical Concerns With Artificial Intelligence

https://www.youtube.com/watch?v=1LyacmzB1Og

After watching the video, answer the following question:
What are the three big ethical concerns with AI?

Brainy Fact

AIACT.IN (also known as the Draft Artificial Intelligence (Development and Regulation Bill, 2023) is India's first private proposal for regulating Artificial Intelligence technologies. It is yet to be passed by the Parliament.

Ai Task

Ethical & Moral Reasoning

Read the article *10 Wonderful Examples of Using Artificial Intelligence (AI) For Good*

https://www.forbes.com/sites/bernardmarr/2020/06/22/10-wonderful-examples-of-using-artificial-intelligence-ai-for-good/

Which example did you like the best?

Why is AI Ethics Important?

Nowadays, we can easily get huge data, and it keeps growing. When Artificial Intelligence (AI) and data work together, they find patterns and help us make decisions, thus, making life easier. But AI also brings challenges and ethical risks that we cannot ignore.

Sometimes, technology focuses only on what is easy to measure, ignoring how people feel. This can give a false sense of security. Technology should respect human values and rights as well as focus on what is best for people. So, there should be rules or guidelines for ethical AI.

When designing AI systems, we should prioritise the well-being of the people who use them considering their mental, physical, and social health.

AI ethics is important for given reasons:

- As Artificial Intelligence becomes more integrated into our lives, and has the potential to impact various aspects of society, including privacy, employment, and decision-making. Ensuring that AI systems are developed and used ethically helps prevent harm to individuals and communities.
- Ethical AI promotes fairness and equity. Without proper guidelines, AI algorithms may unintentionally reinforce biases or discriminate against certain groups of people. Ethical considerations help mitigate these risks and ensure that AI systems treat everyone fairly.
- AI ethics fosters trust and accountability. When people trust that AI technologies are designed and used responsibly, they are more likely to accept and adopt them. Ethical guidelines also hold developers and users accountable for the consequences of AI systems, encouraging responsible behaviour and decision-making.

AI ethics is important for creating a future where Artificial Intelligence benefits society while respecting human values, rights, and well-being.

The Five Pillars of AI Ethics

AI ethics is a multidisciplinary field focused on maximising the positive effects of AI while minimising unintended or negative consequences. The core principles of AI ethics are fairness, robustness, explainability, transparency, and privacy. These areas of focus are essential for ensuring the trustworthiness of AI.

Fairness

Fairness is a crucial aspect of ethics in AI because it ensures that AI systems treat all individuals and groups equitably and without bias. In AI, fairness means that the outcomes produced by algorithms do not disproportionately harm or advantage specific demographics based on characteristics such as race, gender, religion, ethnicity, or socioeconomic status.

Example: A hiring algorithm used by a company must evaluate candidates solely based on their qualifications and experience, without being influenced by gender, race, or age.

Explainability

Explainability in AI is crucial for ensuring that the decisions made by AI systems are understandable to humans. It pertains to the transparency and clarity of AI systems, enabling users to understand the decision-making process and forecasts of algorithms. It facilitates stakeholders in comprehending the fundamental reasoning, elements, and factors influencing algorithmic results, promoting confidence, responsibility, and ethical application of AI technologies. The essence of explainability lies in guaranteeing that AI systems are open, answerable, and in harmony with ethical standards.

Example: Imagine you're using a smart algorithm to determine who gets a loan. If the algorithm denies someone a loan, they have the right to know why. Explainable AI provides this explanation, helping people understand how decisions are made and why certain outcomes occur.

Robustness

Robustness in AI ethics refers to the capacity of AI systems to perform reliably and accurately across various conditions, while minimising unintended consequences and harmful impacts. It is a fundamental aspect of ethical AI because unreliable or biased systems can lead to significant societal harm. Robustness in AI systems implies their ability to deliver precise and dependable outcomes under various circumstances and for longer periods. It ensures that AI algorithms and systems function as intended without facing unforeseen errors or deviating from their designated behaviour.

Example: Self-driving cars need to be robust to operate safely in various weather conditions and road scenarios. A robust AI system in a self-driving car should be able to correctly identify and respond to pedestrians in both clear and foggy conditions.

Transparency

Transparency in AI means being open and clear about how AI systems are created, how they work, and what impacts they might have. It involves providing straightforward information about the data, algorithms, and decision-making processes used in AI applications. This openness encourages accountability, allows for scrutiny, and helps people to make informed choices about the ethical and social implications of AI technologies.

Example: Social media platforms using AI to moderate content should disclose the criteria and algorithms they use to flag or remove posts. If a platform removes content for hate speech, it should clearly explain the rules and the AI's decision-making process to the users.

Privacy

Privacy involves individuals having control over their personal information and avoiding unwarranted interference in their lives. It encompasses the right to keep aspects of one's life private, such as private messages, activities, and data. Privacy is crucial as it safeguards individual autonomy, dignity, and freedom from unnecessary intrusion.

Example: An AI-powered personal assistant, like Siri or Alexa, should ensure that users' voice recordings and personal data are kept secure and not misused. These assistants should not store or share sensitive information without explicit user consent, and data encryption should be used to protect user privacy.

Ethical & Moral Reasoning

Video Session

Watch the video on *AI FOR GOOD - Ethics in AI* and answer the question asked:
https://www.youtube.com/watch?v=vgUWKXVvO9Q

Elaborate on the statement: "We need to choose how we use AI, or else AI will choose how to use us."

AI Reboot

1. Fill in the blanks:
   a. Ethical AI promotes fairness and _____.
   b. Machines learn ethics and bias from _____.
2. Name any 2 pillars of AI Ethics.

   _____

   _____
3. In 2018, Amazon scrapped its AI recruiting tool that showed bias against women. Which ethical AI principle was not being followed?

   _____
4. Define privacy.

   _____

Bias, Bias Awareness, AI Bias and Sources of Bias

A facial recognition algorithm might find it easier to identify a white person compared to a dark complexion person due to the prevalence of white faces in the training data. This discrepancy can unfairly impact individuals from distinct groups, reinforcing inequality and oppression. The challenge lies in the unintentional nature of these biases, which often go unnoticed until they manifest in the software.

Bias

Bias is defined as prejudice against individuals or groups, especially in ways that are considered unfair. "Bias in AI" has long been a key area of research and attention in the machine learning community. It refers to situations where ML-based data analysis systems are biased against certain groups of people. These biases usually reflect the prevailing social biases related to race, gender, biological sex, age, and culture. AI systems learn to make decisions based on training data, which may include biased human decisions or reflect historical or social inequities.

Bias Awareness

In today's connected world, AI technologies are becoming more important in different areas of our lives, such as healthcare, finance, and criminal justice. However, as AI systems become more common, it's crucial to recognise and address the biases they may have. **Bias awareness** means understanding that AI systems can show unfair preferences due to factors like training data used to train the AI models, rules they follow, the algorithms they use, or the principles with which the AI model was designed. This awareness involves understanding that AI may occasionally make biased decisions because of how AI model was developed or trained.

AI Reboot

Answer the given questions:

1. Who comes to your mind when you think of a teacher or cook? Do you think you are biased?

2. What do you mean by bias awareness?

3. AI 's training data can have bias. (State True or False)

AI and Bias

AI bias is a phenomenon that occurs when algorithm results are systematically biased against a certain gender, language, race, wealth, etc. AI bias leads to a skewed output. Algorithms can have inherent biases because AI models are created by individuals with conscious or unconscious preferences, and these preferences, which may not be discovered until the algorithm is used publicly (as we saw in the case of the hand dispenser and Google's photo tagging algorithm).

Bias is one of the biggest challenges facing AI. Although all programmers try to have absolute factual data, there is an inevitable bias when exploring the depth to which AI can be used. An inherent problem with AI systems is that they are as good or as bad as the trained data. Bad data usually carries racial, gender, community, or ethnic biases in algorithms responsible for critical decisions go unrecognized, they can lead to unethical and unfair consequences. In the future, these biases may worsen, especially as many AI recruitment systems will continue to use incorrect data for training. Therefore, there is an urgent need to train these systems with unbiased data and to develop algorithms that are easy to interpret.

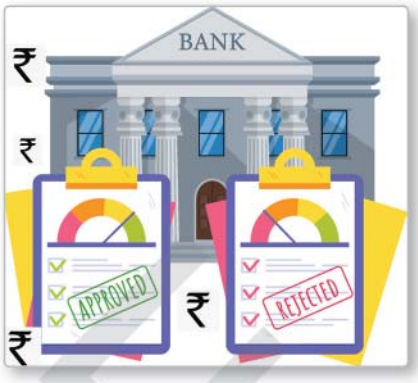
Sources of Bias

Data bias happens when some parts of a dataset are given too much weight or are over-represented implying the dataset doesn't accurately reflect what the machine learning model is meant to do, leading to unfair outcomes and poor accuracy.

Often, biased outcomes often result in discriminate against certain groups of people, such as those based on age, race, culture, or sexual orientation. As AI systems become more common, the risk of data bias is that it can amplify existing discrimination. To prevent any ambiguity or reduce the biasness it is important to identify the source of bias and take necessary steps.

Addressing AI bias involves thorough examination datasets, machine learning algorithms, and other elements of AI systems to identify sources of potential bias.

Training Data Bias



Training data bias occurs when the data used to develop AI systems is unrepresentative, incomplete, or skewed. If the data reflects existing prejudices or excludes certain groups, the AI system may learn and perpetuate these biases.

AI systems make decisions based on data they are trained on, so it is crucial to check datasets for bias. One way to do this is to examine whether certain groups are over- or under-represented in the data. For example, a medical AI system trained on data from male patients may not perform well for female patients, leading to misdiagnoses.

Similarly, an AI system used for loan approvals might be biased if the training data primarily includes applicants from affluent neighbourhoods, disregarding those from poorer areas. Bias can also occur in data labelling.

Algorithmic Bias

Algorithmic bias refers to the bias that may exist in the design, implementation, and outcomes of algorithms used in Artificial Intelligence (AI) systems. This bias can result in unfair or discriminatory outcomes, often reflecting the prejudices or limitations of the data used to train the algorithm, as well as the assumptions and decisions made by developers during the algorithm's creation.

For example, if an AI algorithm used in hiring processes is trained on historical data that reflects biased hiring decisions, such as favouring one demographic group over another, the algorithm may perpetuate these biases when making new hiring recommendations.

Addressing algorithmic bias requires careful consideration of the data used to train algorithms, as well as transparency and accountability in the design and implementation processes. Techniques such as bias detection, fairness-aware algorithms, and diverse data collection can help mitigate algorithmic bias and promote more equitable outcomes in AI systems.

Cognitive Bias

Cognitive bias refers to systematic patterns of deviation from rationality or objectivity in judgment or decision-making. These biases are often influenced by factors such as emotions, personal experiences, and social norms, leading individuals to make judgments or decisions that may not be entirely rational or impartial.

For example, Imagine a person who strongly believes that climate change is not real. When presented with scientific evidence supporting climate change, they might dismiss it or interpret it in a way that aligns with their preconceived notion, while ignoring or downplaying evidence to the contrary. This individual's confirmation bias prevents them from objectively considering added information and may reinforce their existing beliefs, even in the face of contradictory evidence.

Cognitive biases can impact various aspects of life, including personal relationships, business decisions, and societal perceptions. Recognising and understanding these biases is essential for making more informed and rational decisions, both individually and collectively. Strategies such as critical thinking, mindfulness, and seeking diverse perspectives can help mitigate the influence of cognitive biases.

AI Bias in Real Life

AI bias in real life means that the decisions made by AI systems are not always fair or accurate because they are influenced by biases. Just like people, AI systems can make unfair decisions if they are trained on biased data or programmed with biased rules.

Brainy Fact

Tay was an Artificial Intelligence Chatbot that was originally launched by Microsoft Corporation on Twitter on March 23, 2016. When the bot started posting inflammatory and offensive tweets through its Twitter account, it caused a subsequent controversy, prompting Microsoft to shut down the service just 16 hours after its launch. According to Microsoft, this was caused by trolls that "attacked" the service, as the bot's responses were based on its interaction with people on Twitter.

Healthcare

AI model utilised for medical diagnosis can be biased if there is an over-representation of one demographic group in the training data compared to another. For example, an AI system for melanoma detection (a skin disease) trained primarily on fair skinned individuals may not perform well for darker skinned ones, thus leading to misdiagnosis and delayed treatment.

Education

Drafting essays or grading standardised tests with AI can discriminate against test takers based on culture, dialect, or the way of writing. Failure to recognize cultural expressions and languages may unfairly disadvantage students from specific regions.

Finance

Biased training data or attributes may inadvertently lead some AI algorithms used in credit scoring to discriminate against certain demographics. For instance, if historical loan approving data has prejudice lending practices patterns, then AIs trained on such models might initiate them hence making it difficult for lower-level communities to have equal access to loans.

Criminal Justice

If AI tools, trained on a system that had biases during the sentencing of inmates before their trial or parole process, then subsequently it could also be predictive of bias. For instance, because past cases disproportionately target people from certain races or socio-economic classes, this might result into categorising many people from these groups as high risks thereby leading to unfairness in sentencing procedures.

Ai Task

Experiential Learning

Visit the website PortraitAI art generator (https://ai-art.tokyo/en/). Users upload their selfies here, and the Artificial Intelligence uses its understanding of Baroque and Renaissance portraits to draw their portrait in the style of a master. If a person is white, the result is very good. The drawback is that the most popular paintings of this era were of European whites. So, the database consists of mainly white people and an algorithm that relies on the same database tends to make you look 'fairer' while drawing your image.

Mitigating Bias in AI Systems

Artificial Intelligence bias is the result of human biases, as people select the training data for machine learning algorithms and determine how the algorithm's results should be applied. Therefore, people must advocate for ethical AI and find ways to mitigate bias. Data science leaders, policymakers, and other key stakeholders are now prioritising ways to mitigate bias and limit the potentially negative impacts of automated decisions in real-world applications.

There are several reasons to alleviate bias in AI systems. First, when you have a bias in your AI system, you amplify whatever problems you have. Unfairness and discrimination are bad on their own, and when you add an AI system on top of them, they get worse. For instance, using biased algorithms for hiring can disadvantage certain groups and perpetuate systemic discrimination. Second, when AI systems are biased, people lose trust in technology. If you don't trust your AI system to behave fairly towards you, you may choose not to interact with it at all. And this can be detrimental not only to you but to everyone. Finally, we strive to tackle bias because it's inherent to certain ethical principles and we want to ensure AI systems are developed and used responsibly.

Strategies for Mitigating Bias

There are several strategies and techniques for mitigating bias in AI systems:

- **Ethical Guidelines and Policies:** Formulate ethical guidelines and policies to ensure fairness and mitigate biases in decision-making. Always keep provisions for regular review and update of guidelines to address new issues.
- **Diverse Data:** To reduce bias, we should use lots of different kinds of information to teach AI. So, AI can learn from many different examples and viewpoints, making it less likely to be biased.
- **Data Transparency and Quality:** By maintaining transparency of the data inputs that are fed into the decision-making process and maintaining data quality by continuously checking data sources for potential biases.
- **Bias Testing and Auditing:** Regularly conduct testing and auditing to identify biases in algorithms, models, or decision-making processes. This should be done regularly, and systems updated to mitigate biases detected.
- **Algorithmic Fairness**: We can make AI systems fairer by using special algorithms that are designed to be fair. These algorithms make sure to consider fairness when making decisions, helping to reduce bias.
- **Accountability and Transparency:** Ensure accountability of decision makers for their decisions and potential biases.
- **Feedback Mechanisms:** Create feedback mechanisms that allow stakeholders to flag potential biases in the decision-making process. Actively seek and consider feedback from stakeholders to mitigate biases and ensure fairness and inclusiveness.
- **Continuous Improvement:** Regularly evaluate decision-making processes for biases and establish mechanisms for continuous improvement based on feedback, data, and emerging research.

Developing AI Policies

Creating rules for AI is crucial to ensure its ethical and fair use. Clear guidelines are necessary regarding the deployment of AI, with consideration given to everyone's input in the rule-making process. Before using AI, we should check for any problems and have plans to fix them.

Fair Treatment: Make rules to ensure AI treats everyone fairly and respects their rights.

Transparency: Provide honest explanations of how AI functions and ensure transparency in its decision-making processes.

Safety: Implement measures to keep AI systems safe and secure for users.

Accountability: Hold decision-makers responsible if something goes wrong with AI usage.

Clear Standards: Establish clear rules and standards for using AI, covering areas like data protection and bias prevention.

Inclusivity: Consult diverse stakeholders, including government officials, businesses, scientists, and community groups, when creating AI policies.

Risk Assessment: Before using AI, identify potential problems and risks, and have plans in place to address them.

Understanding the components of AI policies involves examining guidelines and principles established by various organizations and regulatory bodies.

IBM AI Ethics Board

Focus: Ethical development and deployment of AI technologies across various industries.

Components:

- Principles for responsible AI development and deployment, including fairness, explainability, transparency, privacy, and robustness.
- Drafting rules and guidelines that people should adhere for using and preparing AI stuff.
- Identifying how to be fair, honest, and responsible and reduce unfairness in AI stuff.
- Discussing with all the people involved in AI, like researchers, government people, and businesses, to work together on being ethical with AI.
- Educating people about AI ethics through education and resources so everyone knows how to do AI stuff the right way.

Microsoft's Responsible AI Page

Focus: Corporate responsibility and ethics in AI

Components:

- Guidelines for making and using AI responsibly, like being fair, dependable, respect of privacy, and welcoming everyone.
- Helpful stuff like fairness checks and programs that find biases, to make sure ethical concerns are part of AI projects.
- Real-life examples and good ways to use AI responsibly in different areas, showing what works well.

Artificial Intelligence at Google

Focus: Corporate AI ethics and governance

Components:

- Google's rules for making AI ethically, covers fairness, safety, privacy, and accountability.
- Ways to make AI systems that focus on what's important to people and making society better.
- Promises to be open, work together, and keep getting better at how AI is controlled and decisions are made.

European Union's Ethics Guidelines for Trustworthy AI

Focus: Ethical guidelines for AI development and deployment in the EU

Components:

- Rules for AI you can trust, like letting people make their own choices, avoiding harm, being fair, and taking responsibility.
- Needs for AI to be clear, understandable, and able to be checked in detail.
- Ideas for making sure people are incharge and there are ways to check and take responsibility for AI that affects society a lot.

Brainy Fact

IBM's AI Fairness 360 is an open-source toolkit designed to address bias in machine learning models. It includes over 70 fairness metrics to help users detect bias, indicating its robust capability for identifying potential bias sources. The toolkit also offers more than 10 algorithms for mitigating bias, such as optimising the preprocessing stage, prejudice remover, and regular algorithms. With its comprehensive features, educational resources, and validation mechanisms, AI Fairness 360 aims to promote fairness and equity in AI applications.

Building Trusted AI pipelines

(Using Open-Source)

AI and Ethical Concerns

Ethics is defined as the discipline that deals with human moral obligations and duties. It is a set of ethical principles that regulate the behaviour and actions of individuals or groups. Artificial Intelligence ethics is a part of ethical technology specific to robots and other artificially intelligent machines. Several ethical concerns regarding AI include:

- If AI produces human-like results, can it also take human-like decisions?
- If AI decides whether to grant bank loans, will the AI algorithms be fair and just?
- If AI makes human decisions, is it also worthy of human confidence?
- AI is basically data + mathematical model + data-based training + prediction. What if the data provided for training is unintentionally skewed or biased?
- Will AI lead to significant job displacement?

These questions and concerns come up all the time, so the "ethics of Artificial Intelligence" is very important. We must understand how Artificial Intelligence works and the framework of Artificial Intelligence ethics.

Ethical Dilemma

An ethical dilemma is a situation in which a person faces a choice between conflicting moral principles or values. It often involves tough decisions where there are competing interests or where doing what is considered right may result

in undesirable outcomes. Ethical dilemmas can arise in various contexts, such as in personal relationships, professional settings, or societal issues. Resolving ethical dilemmas requires thoughtful consideration of the consequences of different actions and balancing conflicting ethical concerns.

Example:

- **Scenario**: You work for a pharmaceutical company developing a new drug to treat a rare disease. During clinical trials, it becomes evident that the drug is effective in treating the disease, but it also has significant side effects in a small percentage of patients. The company is under pressure to release the drug quickly due to the urgent need for treatment, but there are concerns about the potential harm caused by the side effects.
- **Ethical Dilemma**: On one hand, releasing the drug could provide relief to patients suffering from the rare disease, potentially saving lives, and improving quality of life. On the other hand, there's a risk of causing harm to patients due to the side effects, which could lead to serious health complications or even fatalities.

Ai Reboot

1. Fill in the blank

   _____ is one of the biggest challenges facing AI.

2. A data scientist favours datasets gathered from Indians rather than sampling from a range of populations around the globe. Name this type of bias.

   _____

3. ChatGPT falsely created a sexual harassment scandal involving a law professor, claiming he was accused of making inappropriate comments and touching a student during a class trip to Alaska. An email informed the professor that ChatGPT had listed him in a Washington Post article. However, upon investigation, it was found that the article did not exist, and the professor had never taken such a trip or faced such accusations. This incident highlights how easily AI can generate convincing falsehoods, raising awareness about the limitations of AI technology and the need for caution in believing potentially damaging fabrications. What was the source of this bias?

   _____

4. Name any 2 steps that should be taken to mitigate the above bias.

   _____

   _____

5. What is an ethical dilemma?

   _____

Ethical & Moral Reasoning

Video Session

Scan the QR code or go to the link to watch the video and answer the questions given below: Humans Need Not Apply

https://www.youtube.com/watch?v=7Pq-S557XQU

What have you learned from this video?

_____

_____

Moral Machine

With the rapid development of Artificial Intelligence comes concerns about how machines will make moral decisions, and the major challenge of quantifying societal expectations about the ethical principles that should guide machine behaviour. To address this challenge, an online experimental platform is designed to explore the moral dilemmas faced by autonomous vehicles called **Moral Machines.**

It was developed by researchers at the Massachusetts Institute of Technology (MIT). It is an online platform designed to explore ethical dilemmas in AI through interactive decision-making scenarios. It is a platform for gathering a human perspective on moral decisions made by machine intelligence, such as self-driving cars. It generates moral dilemmas, where a driverless car must choose the lesser of two evils, such as killing two passengers or five pedestrians. As an outside observer, people judge which outcome they think is more acceptable. They can then see how their responses as compared with other people. If they are feeling creative, people can also design their own scenarios for others to view, share, and discuss.

The Moral Machine shows you different situations that make you think about what's right and wrong. For instance, picture yourself driving a self-driving car. Suddenly, it faces a tough choice: should it swerve to avoid hitting people on the street, even if it puts the passengers in danger, or stay on track and risk hurting those outside? What would you do? And why?

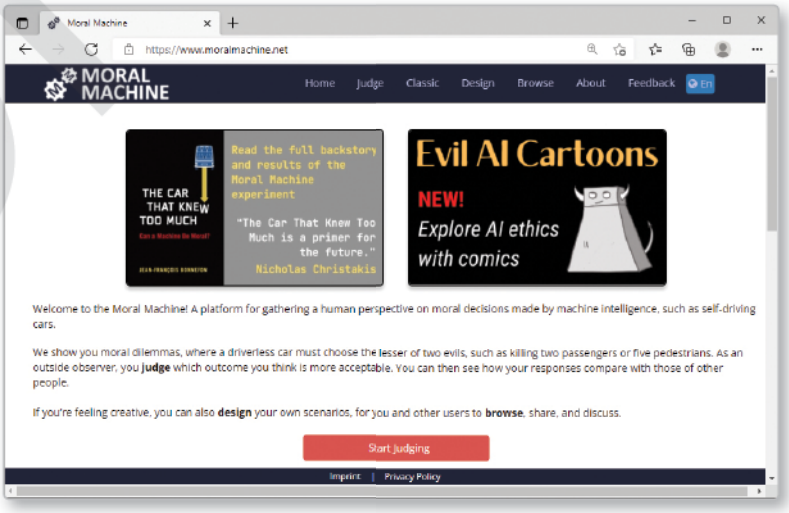
The situations on the Moral Machine might not be real, but they mirror actual ethical problems that AI makers, government people, and everyone else need to think about. As AI gets better and more common, it's important that we talk about the right and wrong ways to use it. The Moral Machine helps start those conversations, making people more aware and ethical when it comes to AI.

Experiential Learning

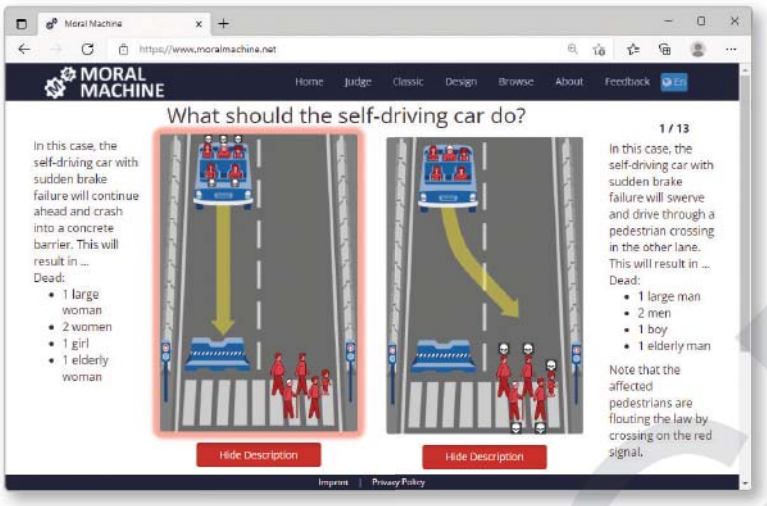
AI GAME 01 — Moral Machine

This platform is based on Machine Learning which shows the machine intelligence in making moral decisions. Perform the following steps:

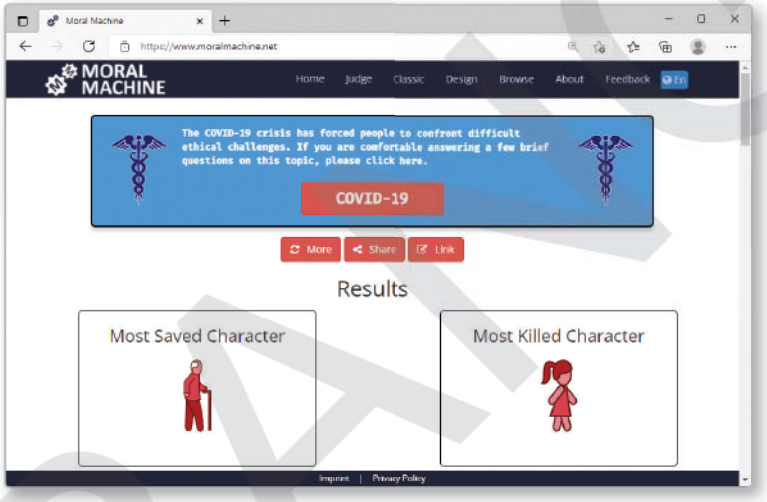
Step 1: Visit the following link and click on the Start Judging button:
https://www.moralmachine.net/



Step 2: Read the description and select the appropriate action according to your choice.



Once you have completed all the 13 situations, you will get the result:



Video Session

Ethical & Moral Reasoning

Scan the QR code or visit the following link to watch the video: What if we could use machine intelligence to save more lives during a natural disaster?

https://www.youtube.com/watch?v=Trv_5dWS4BY

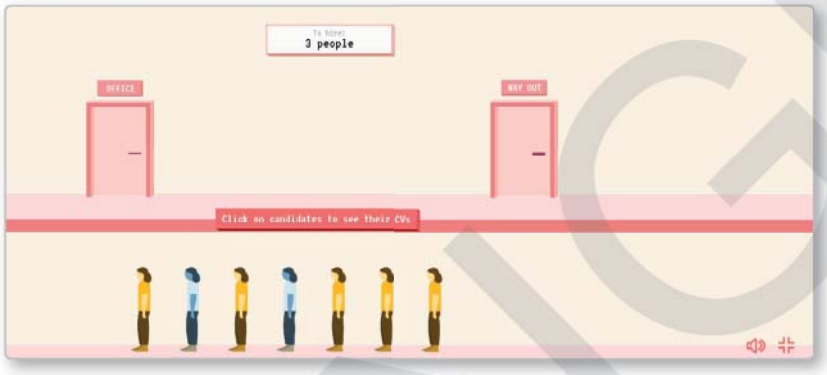
After watching the video, answer the following question:

Find 2-3 more such examples.

AI GAME 02 Survival of the Best Fit

Survival of the Best Fit is a game that teaches about bias in AI hiring. It shows how using AI wrongly can make machines have the same biases as people, leading to more unfairness. Many discussions about AI see it as a danger to humans, but this game shifts that perspective. It helps people learn about AI and ask for more responsibility from those making these widespread software systems.

Play the bias game:- https://www.survivalofthebestfit.com/game/



After playing the game, answer the following questions:

a. Do you think you were biased while hiring people?

b. How can you make your decisions more transparent?

At a Glance

- AI powers navigation apps, helping you find the best or most eco-friendly routes.
- Artificial intelligence ethics is a set of moral principles and practices designed to support the development and responsible use of AI technology.
- AI ethics is a multidisciplinary field focused on maximising the positive effects of AI while minimising unintended or negative consequences.
- Fairness is a crucial aspect of ethics in AI because it ensures that AI systems treat all individuals and groups equitably and without bias.
- Explainability in AI is crucial for ensuring that the decisions made by AI systems are understandable to humans.
- Robustness in AI ethics refers to the capacity of AI systems to perform reliably and accurately across various conditions.
- Transparency in AI means being open and clear about how AI systems are created, how they work, and what impacts they might have.

- Privacy involves individuals having control over their personal information and avoiding unwarranted interference in their lives.
- Bias is defined as prejudice against individuals or groups, especially in ways that are considered unfair.
- AI bias is a phenomenon that occurs when algorithm results are systematically biased against a certain gender, language, race, wealth, etc.
- Algorithmic bias refers to the bias that may exists in the design, implementation, and outcomes of algorithms used in Artificial Intelligence (AI) systems.
- Cognitive bias refers to systematic patterns of deviation from rationality or objectivity in judgment or decision-making.
- If AI tools, were trained on a system that had biases during the sentencing of inmates before their trial or parole process, then subsequently it could also be predictive of bias.
- Ethics is defined as the discipline that deals with human moral obligations and duties.
- An ethical dilemma is a situation in which a person faces a choice between conflicting moral principles or values.
- The Moral Machine game shows you different situations that make you think about what's right and wrong. For instance, picture yourself driving a self-driving car.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. What is the primary reason for mitigating bias in AI systems?
   a. To amplify existing problems ◯ b. To reinforce societal inequalities ◯
   c. To enhance trust and reliability ◯ d. To perpetuate systemic discrimination ◯
2. Which of the following is NOT one of the core principles of AI ethics?
   a. Fairness ◯ b. Explainability ◯
   c. Cost-effectiveness ◯ d. Privacy ◯
3. Robustness in AI systems refers to:
   a. The ability to explain decisions ◯ b. Reliability and accuracy across various conditions ◯
   c. Transparency in data usage ◯ d. Cost reduction ◯
4. Transparency in AI involves:
   a. Keeping algorithms secret ◯
   b. Openness about data, algorithms, and decision-making processes ◯
   c. Reducing the computational cost ◯
   d. Hiding the decision-making process from users ◯

5. Why is privacy important in AI?

a. It helps in faster decision-making. ◯ b. It protects individual independence and dignity. ◯

c. It reduces the cost of data storage. ◯ d. It makes AI systems more accurate. ◯

6. What is one potential consequence of biased AI systems mentioned in the text?

a. Decreased trust in technology ◯ b. Increased fairness in decision-making ◯

c. Enhanced user satisfaction ◯ d. Greater reliance on AI systems ◯

7. What is the primary cause of facial recognition algorithms having difficulty identifying people with darker complexions?

a. Poor camera quality ◯ b. Lack of diverse training data ◯

c. Over-representation of older faces ◯ d. Use of outdated algorithms ◯

8. Which of the following is NOT a strategy for mitigating bias in AI systems?

a. Ethical guidelines and policies ◯ b. Unsupervised learning techniques ◯

c. Data transparency and quality ◯ d. Bias testing and auditing ◯

9. What does bias in AI typically reflect?

a. Technological advancements ◯

b. Social prejudices about race, gender, age, and culture ◯

c. Random data anomalies ◯

d. Errors in coding ◯

10. What is meant by bias awareness in the context of AI?

a. Understanding that AI systems can show unfair preferences ◯

b. Ensuring AI systems are technologically advanced ◯

c. Using only the latest AI algorithms ◯

d. Avoiding the use of AI in sensitive areas ◯

B. Fill in the blanks.

1. Bias is defined as _____ against individuals or groups, especially in ways that are considered unfair.
2. AI systems learn to make decisions based on _____ data.
3. Data bias happens when some parts of a _____ are given too much weight or are over-represented.
4. _____ AI promotes fairness and equity.
5. _____ ensures that AI algorithms and systems function as intended without facing unforeseen errors or deviating from their designated behaviour.
6. The Moral Machine is an online platform designed to explore _____ dilemmas in AI through interactive scenarios.
7. _____ is defined as the discipline that deals with human moral obligations and duties.
8. _____ refers to the bias that can be present in the rationality, implementation, and outcomes of algorithms.
9. _____ refers to systematic patterns of deviation from rationality or objectivity in judgement or decision-making.
10. AI bias in real life means that the decisions made by AI systems are not always _____ or accurate because they are influenced by biases.

C. State whether the following statement is True or False.

1. Tay was an Artificial Intelligence Chatbot that was originally launched by Microsoft Corporation on Twitter on March 23, 2016. _____
2. Ethical dilemmas only occur in personal relationships and not in professional or societal contexts. _____
3. Bias in AI systems can lead to unfair outcomes but doesn't affect societal perceptions. _____
4. AI can only cause physical harm, not any other type of harm. _____
5. Robustness in AI ensures that AI systems function as intended without facing unforeseen errors. _____
6. The Moral Machine is an offline experimental platform designed to explore ethical dilemmas in AI. _____
7. AI systems can make unfair decisions if they are trained on biased data or programmed with biased rules. _____
8. Algorithmic bias only occurs during the data collection phase of AI development. _____
9. Bias in AI is always easy to detect and eliminate before the AI system is publicly used. _____
10. Artificial Intelligence is the product of human imagination. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. What is AI ethics?

Ans. AI ethics refers to a set of moral principles and practices designed to guide the development and responsible use of Artificial Intelligence technology.

2. Why is privacy a significant concern in AI ethics?

Ans. Privacy protects individual independence, dignity, and the right to be free from unnecessary intrusion. In AI, it involves ensuring that personal data and activities are kept secure and not misused.

3. What is AI bias?

Ans. AI bias is a phenomenon which occurs when algorithm results are systematically biased against a certain gender, language, race, wealth, etc.

4. List any 2 strategies and techniques for mitigating bias in AI systems.

Ans. Two strategies and techniques for mitigating bias in AI systems are:

- Ethical guidelines and policies
- Data transparency and quality

5. Explain the term "Bias Awareness".

Ans. Bias awareness means understanding that AI systems can show unfair preferences due to factors like training data used to train the AI models, the rules they follow, the algorithms they use, or the principles with which the AI model was designed.

6. How can programmers avoid bias?

Ans. To avoid bias, programmers can follow the philosophical basis of asking questions out of curiosity, seeking to transcend their own perspectives, and learning to see the world outside their own biases.

7. How can training data bias affect AI systems?

Ans. Training data bias occurs when the data used to train AI systems is unrepresentative or incomplete. This can cause the AI to perform poorly for underrepresented groups, leading to unfair outcomes and inaccuracies.

8. What is cognitive bias, and how does it impact decision-making?

Ans. Cognitive bias refers to systematic patterns of deviation from rationality in judgement or decision-making, influenced by emotions, personal experiences, and social norms. It impacts decision-making by causing individuals to make irrational or partial judgements.

B. Long answer type questions.

1. Explain the five pillars of AI ethics.

Ans. The five pillars of AI ethics are fairness, robustness, explainability, transparency, and privacy. The description of each are as follows:

- **Fairness:** Fairness is a crucial aspect of ethics in AI because it ensures that AI systems treat all individuals and groups equitably and without bias. In AI, fairness means that the outcomes produced by algorithms do not disproportionately harm or advantage specific demographics based on characteristics such as race, gender, religion, ethnicity, or socioeconomic status.
- **Explainability:** Explainability in AI is crucial for ensuring that the decisions made by AI systems are understandable to humans. It pertains to the transparency and clarity of AI systems, enabling users to understand the decision-making process and forecasts of algorithms.
- **Robustness:** Robustness in AI ethics refers to the capacity of AI systems to perform reliably and accurately across various conditions, while also minimising unintended consequences and harmful impacts. It is a fundamental aspect of ethical AI because unreliable or biased systems can lead to significant societal harm.
- **Transparency:** Transparency in AI means being open and clear about how AI systems are created, how they work, and what impacts they might have. It involves providing straightforward information about the data, algorithms, and decision-making processes used in AI applications. This openness encourages accountability, allows for scrutiny, and helps people make informed choices about the ethical and social implications of AI technologies.
- **Privacy:** Privacy involves individuals having control over their personal information and avoiding unwarranted interference in their lives. It encompasses the right to keep aspects of one's life private, such as private messages, activities, and data. Privacy is crucial as it safeguards individual autonomy, dignity, and freedom from unnecessary intrusion.

2. Discuss the concept of the ethical dilemma with an example.

Ans. An ethical dilemma is a situation in which a person faces a choice between conflicting moral principles or values. It often involves tough decisions where there are competing interests or where doing what is considered right may result in undesirable outcomes. Ethical dilemmas can arise in various contexts, such as in personal relationships, professional settings, or societal issues. Resolving ethical dilemmas requires thoughtful consideration of the consequences of different actions and balancing conflicting ethical concerns.

Let us understand the concept of ethical dilemma with the help of an example.

Scenario: You work for a pharmaceutical company developing a new drug to treat a rare disease. During clinical trials, it becomes evident that the drug is effective in treating the disease, but it also has significant side effects in a small percentage of patients. The company is under pressure to release the drug quickly due to the urgent need for treatment, but there are concerns about the potential harm caused by the side effects.

Ethical Dilemma: On one hand, releasing the drug could provide relief to patients suffering from the rare disease, potentially saving lives, and improving quality of life. On the other hand, there's a risk of causing harm to patients due to the side effects, which could lead to serious health complications or even fatalities.

3. Explain the different sources of bias in AI systems and how they can lead to unfair outcomes.

Ans. Bias in AI systems can stem from several sources, including training data bias, algorithmic bias, and cognitive bias.

- **Training data bias:** This occurs when the data used to train AI systems is unrepresentative, incomplete, or skewed. For instance, if a medical AI system is trained primarily on data from male patients, it may not perform well for female patients, leading to misdiagnoses. Similarly, an AI used for loan approvals might be biased if it primarily includes applicants from affluent neighbourhoods, thereby ignoring applicants from poorer areas.
- **Algorithmic bias:** This type of bias arises during the design and implementation of algorithms. If an AI hiring algorithm is trained on historical data that reflects biased hiring decisions, such as favouring one demographic group over another, the algorithm may perpetuate these biases in new hiring recommendations.

- **Cognitive bias:** This refers to systematic patterns of deviation from rationality or objectivity in judgement, influenced by factors like emotions and personal experiences. For example, a person who strongly believes that climate change is not real might dismiss scientific evidence supporting it, thus reinforcing their existing beliefs. Cognitive biases can lead to irrational or partial judgements, impacting AI development and application.

Each of these biases can result in AI systems that unfairly discriminate against certain groups, leading to unethical and unfair outcomes in various sectors such as healthcare, finance, and criminal justice.

C. Competency-based/Application-based questions:

1. The below question consists of two statements: Assertion (A) and Reasoning (R).

   Assertion(A): Transparency and explainability are crucial in AI systems to promote trust and accountability.

   Reasoning(R): Transparency ensures that users understand how AI systems make decisions, while explainability provides insight into the reasoning behind those decisions. This promotes trust by allowing users to verify the fairness and reliability of AI systems. Additionally, accountability is enhanced when the decision-making process is transparent, as it enables stakeholders to identify and address any biases or errors in the AI system.

   Answer the question by selecting the appropriate option given below:

   a. Both A and R are true and R is the correct explanation of A
   b. Both A and R are true and R is not the correct explanation of A
   c. A is true but R is false
   d. A is false but R is true

Ans. a. Both A and R are true and R is the correct explanation of A

2. Aditi, a young entrepreneur, applies for a business loan to expand her small enterprise. She has an excellent credit score, a solid business plan, and a history of successful repayments. However, the bank she approaches uses an AI system to evaluate loan applications. This AI system was trained primarily on data from affluent urban areas and does not consider factors relevant to Aditi's rural context.

   Despite her strong qualifications, the AI system flags her application as high risk because the training data does not adequately represent rural entrepreneurs. As a result, Aditi's loan application is denied, and she is unable to expand her business. What are the potential consequences of using biased AI systems in financial services, and how can such biases be addressed to ensure fair treatment of all applicants?

Ans. Using biased AI systems in financial services can lead to unfair treatment of certain groups, such as rural entrepreneurs like: Aditi. This can result in qualified applicants being denied opportunities, perpetuating economic disparities. To address such biases:

- **Inclusive data collection:** Ensure the training data includes diverse representations from various demographics, including rural areas. This helps the AI system to learn patterns that are applicable to a broader range of applicants.
- **Bias detection and correction:** Implement algorithms that detect and correct biases during the evaluation process. This can involve techniques such as fairness constraints and bias mitigation strategies.
- **Human oversight:** Complement AI assessments with human judgement to catch and mitigate unfair decisions. Loan officers can review flagged applications to ensure qualified candidates are not unjustly denied.
- **Transparency:** Make the AI decision-making process transparent so applicants understand why their applications are accepted or denied. Providing clear reasons for decisions can help build trust and allow applicants to address any specific issues.

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. Why is bias in AI considered one of the biggest challenges?
   a. AI systems can only be used for specific tasks
   b. Bias can lead to unethical and unfair consequences
   c. AI development is too expensive
   d. AI systems are difficult to interpret

2. What is training data bias?
   a. Bias caused by incorrect programming
   b. Bias due to unrepresentative or incomplete training data
   c. Bias from user interaction with AI
   d. Bias occurring during the deployment of AI systems

3. What can be a result of algorithmic bias in AI systems?
   a. Fair and equal treatment of all users
   b. Increased efficiency and accuracy
   c. Unfair or discriminatory outcomes
   d. Enhanced user satisfaction

4. What is cognitive bias in AI?
   a. The result of biased data inputs
   b. Bias caused by the emotions and personal experiences of individuals
   c. Bias due to technical errors in AI systems
   d. Bias resulting from incorrect labeling of data

5. What is an example of AI bias in the healthcare sector?
   a. AI diagnosing all patients equally
   b. AI models performing well on diverse demographic groups
   c. AI failing to diagnose dark-skinned individuals correctly due to training data bias
   d. AI eliminating all forms of bias in medical treatments

6. What can be the impact of AI bias in the financial sector?
   a. Equitable access to loans for all demographics
   b. Increased financial inclusion for low-income communities
   c. Discrimination against certain demographics in credit scoring
   d. Elimination of all prejudices in loan approvals

7. What is the focus of AI ethics as a multidisciplinary field?
   a. Maximizing profits through AI applications
   b. Minimizing legal implications of AI misuse
   c. Maximizing positive effects while minimizing negative consequences
   d. Minimizing the use of AI in everyday life

8. Why is fairness critical in the context of AI ethics?
   a. To accelerate the development of AI technologies
   b. To ensure equal distribution of AI resources
   c. To maximize AI's impact on environmental sustainability
   d. To prevent bias and ensure equitable treatment of individuals

9. What is the purpose of the Moral Machine?
   a. To develop autonomous vehicles
   b. To explore ethical dilemmas in AI
   c. To create interactive decision-making scenarios
   d, To quantify societal expectations about AI ethics

10. How do AI and data collaboration benefit decision-making?
   a. By replacing human judgment entirely
   b. By automating all decision-making processes
   c. By identifying patterns and providing insights
   d. By increasing the complexity of decision-making tasks

B. Fill in the blanks.

1. AI improves _____ forecasting, helping people prepare for major weather events.
2. The harm from AI isn't always _____; it can be less obvious, like unfairness or discrimination.
3. _____ is a crucial aspect of ethics in AI because it ensures that AI systems treat all individuals and groups equitably and without bias.
4. _____ in AI is crucial for ensuring that the decisions made by AI systems are understandable to humans.
5. _____ in AI means being open and clear about how AI systems are created, how they work, and what effects they might have.
6. _____ involves individuals having control over their personal information and avoiding unwarranted interference in their lives.
7. _____ in AI ethics refers to the capacity of AI systems to perform reliably and accurately across various conditions, while minimising unintended consequences.
8. AI systems learn to make decisions based on _____ data, which may include biased human decisions or reflect historical or social inequities.

C. State whether the following statement is True or False.

1. AI systems are incapable of making ethical decisions. _____
2. An ethical dilemma arises when there is no conflict between moral principles or values. _____
3. The Moral Machine is a platform designed to explore ethical dilemmas in AI through interactive decision-making scenarios. _____
4. Bias testing and auditing should be conducted irregularly to identify biases in AI systems. _____
5. Training data bias can result in AI systems that do not perform well for underrepresented groups. _____
6. AI bias can lead to both unethical and unfair consequences in decision-making processes. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. Why is robustness important in AI ethics?
2. What measures can be taken to ensure fair treatment by AI systems?
3. What are the main goal of AI ethics as a multidisciplinary field?
4. Provide an example of AI bias in the healthcare sector.
5. Why is bias awareness important in AI technologies?

B. Long answer type questions.

1. Explain cognitive bias and how it can influence decision-making.
2. How do games like the Moral Machine and Survival of the Best Fit help individuals understand ethical issues and biases in AI?
3. Why is transparency important in AI, and how can it be achieved in practical applications?
4. What is the purpose of intelligence ethics in the development and use of AI technology?
5. What measures can be taken to mitigate the effects of bias in AI systems?

C. Competency-based/Application-based questions:

1. The following question consists of two statements: Assertion (A) and Reasoning (R).

   Assertion(A): Bias in AI systems can lead to unfair outcomes and perpetuate social inequalities.

   Reasoning(R): AI systems make decisions based on the data they are trained on, and if this data reflects societal biases or prejudices, the AI may replicate and amplify these biases in its decision-making process. This can result in unfair treatment of certain groups, exacerbating existing inequalities in areas such as healthcare, finance, and criminal justice.

   Answer the question by selecting the appropriate option given below:

   a. Both A and R are true and R is the correct explanation of A
   b. Both A and R are true and R is not the correct explanation of A
   c. A is true but R is false
   d. A is false but R is true

2. Rahul works for a company that uses autonomous drones to deliver packages in urban areas. One day, a drone encounters a malfunction while navigating through a busy city street. The drone's AI must choose between making an emergency landing in a park, where children are playing, or on a deserted road, risking damage to the drone and its package.

   The drone, programmed to minimize human harm, attempts to land on the deserted road but crashes into a parked car, causing significant damage. The car owner, upset by the incident, demands accountability from the drone company.

   In the case of the drone's malfunction, who should be held accountable for the damage, and what ethical considerations should guide the development of autonomous systems to handle such scenarios?

Ai In Life

Interdisciplinary

Read the following real life examples of AI impacting lives positively!

1. **My Drawings Speak Up (Palestine)**

A young student saw a friend being bullied at school but was afraid to speak out. She realized that a student's drawing can often express what he or she is experiencing in life.

Therefore, she developed a solution to identify and support students who have experienced violence and bullying. She and her mother teamed up to create 'My Drawings Speak Up', an app that utilises machine learning to examine drawings to determine if a child is experiencing violence.

(My Drawings Speak Up: https://www.youtube.com/watch?v=hRTorWpfZsU)

2. **Emotion Detector For Angry Players (Spain)**

Have you noticed that while playing video games, your emotions run free? You most probably have become angry and aggressive. A group of Spanish middle school boys who call themselves 'Happy Llamas' realised that even after the game is over, the negative emotions they experience while playing video games continue to hurt their emotional state. Therefore, they developed 'Emotion Detector for Angry Players', a mobile application that uses image gesture recognition to indicate when players experience healthy (calm) or unhealthy (angry) emotions. When the app detects any emotion, it notifies the player by displaying the shades between green (calm) and red (anger). This indicator helps the players regulate their emotions while playing. Find out 2 more such examples.

Ai Deep Thinking

Creativity and Innovativeness

Read the article - Can a machine learn morality?

https://economictimes.indiatimes.com/news/science/can-a-machine-learn-morality/articleshow/87818790.cms?from=mdr

Then visit Ask Delphi - https://delphi.allenai.org/

Discuss among students your views on Delphi. The following questions may help you frame your views:

- Do you feel this AI has still a lot to learn? Why/Why not?
- If your answer to the above question is 'Yes', what kind of data should be added to Delphi's AI model?
- Do you think such moral machines should be integrated in our justice system? Why/Why not?

Ai Lab

Experiential Learning

1. Play the following game on Kahoot:
   https://create.kahoot.it/share/thought-experiment-trolley-problem-ai-in-autonomous-vehicles/3bf7c44a-f675-491b-a5d0-b0ad38f04d00
2. Play the Facework game! Understand what it means to see like a machine and how machines might fail. If you have a smiling face that matches the machine's data then you could be hired as a babysitter otherwise get ready to become a prison guard! Go to https://facework.app/
3. Have a job interview with an AI HR manager. "An Interview with ALEX" is a 12-minute guided, interactive experience in the browser that immerses the spectator in a job interview with ALEX, a sophisticated artificial intelligence HR hired by a hypothetical tech giant known as "Open Mind Playground." The experience reveals how companies can use machine learning to conceal the true purpose of a job, extract the maximum amount of labour from its employees, and censor information under the radar—by gamifying the work process, applying filters to employees' information intake, and generating customized distractions. Go to
   https://carriesijiawang.com/interview/#two
4. Understand the future of automated media manipulation by either uploading your own photos or trying to detect fake videos. Deep Angel - image manipulation technology (http://deepangel.media.mit.edu/)
5. Check out the gender bias in Google Translate. Translate the following statements to Hindi and check the bias
   a. The nurse is eating.
   b. The doctor is walking.
   c. The mathematician is working.

   You can read more about this at https://scroll.in/article/991275/google-translate-is-sexist-and-it-needs-a-little-gender-sensitivity-training

Class Activity

Collaboration & Teamwork

Divide the class into groups of 3 students each. **CBSE Handbook**

Ask students to bring any printed media like news articles, advertisements, or social media posts, and identify instances of bias based on factors like race, gender, or socio-economic status.

Note down three points on how bias can influence perceptions and stereotypes.

Answers

Exercise (Section A)

A. 1. c 2. c 3. b 4. b 5. b 6. a 7. b 8. b 9. b 10. a

B. 1. Prejudice 2. Training 3. Dataset 4. Ethical 5. Robustness
6. Ethical 7. Ethics 8. Design 9. Cognitive bias 10. fair

C. 1. True 2. False 3. False 4. False 5. True 6. False
7. True 8. False 9. False 10. True

Activity: Role Play **CBSE Handbook**

1. Share the following examples of biased AI systems and their potential consequences and ask students to do a role play to present each scenario:

Social Media Algorithms:

- **Example:** Algorithms on social media platforms often prioritize content based on user engagement metrics, leading to echo chambers and reinforcing existing biases and opinions.
- **Consequences:** Biased social media algorithms can contribute to polarization and the spread of misinformation, undermining democratic discourse and social cohesion.

Loan Approval Systems:

- **Example:** Automated loan approval systems used by banks may use biased criteria that favour certain demographic groups over others, leading to disparities in access to credit.
- **Consequences:** Biased loan approval algorithms can perpetuate financial exclusion and deepen economic inequality, particularly for marginalized communities who may already face barriers to financial resources.

Criminal Sentencing Algorithms:

- **Example:** Some jurisdictions use algorithms to assist judges in determining sentencing decisions based on factors such as criminal history and offense severity. However, these algorithms have been found to exhibit racial biases.
- **Consequences:** Biased sentencing algorithms may result in harsher punishments for individuals from minority groups, contributing to mass incarceration and perpetuating systemic injustice within the criminal justice system.

School Admissions Algorithms:

- **Example:** Educational institutions may use algorithms to assist in the admissions process, considering factors like academic performance and extracurricular activities. However, these algorithms may unintentionally disadvantage students from underprivileged backgrounds.
- **Consequences:** Biased admissions algorithms can limit educational opportunities for disadvantaged students, widening the achievement gap and hindering social mobility.

ARTIFICIAL INTELLIGENCE

CODE 843 | Skill Education

Supplement Based on CBSE Curriculum 2024-25

ORANGE
@orangeeducation

Orange Education
Adding Value to 'Education'
9, Daryaganj, New Delhi-110002
Phone: +91-11-43776600
E-mail: info@orangeeducation.in
Website: www.orangeeducation.in