ARTIFICIAL INTELLIGENCE

Beta

ORANGE

Answer Key

Artificial Intelligence

Part-B: Subject Specific Skills

1. Artificial Intelligence for Everyone

Exercise

🔒 **Unsolved Questions**

SECTION A (Objective Type Questions)

Δi Quiz

A. 1. c   2. b   3. a   4. a   5. b

B. 1. numbers   2. trial and error   3. unsupervised

     4. Deep learning   5. video, images

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

SECTION B (Subjective Type Questions)

A. 1. Natural Language Processing (NLP) is a field of AI that focuses on enabling machines to understand, interpret, and respond to human languages. NLP combines linguistics and machine learning to perform tasks like text analysis, speech recognition, language translation, and sentiment analysis. It is used in applications like chatbots, virtual assistants (e.g., Siri, Alexa), and spam filters.

     2. ● AI can do:
   - o Recognize patterns in large datasets (e.g., image recognition).
   - o Automate repetitive tasks (e.g., chatbots and customer service).
   - ● AI cannot do:
   - o Exhibit genuine creativity like humans (e.g., creating original art from scratch).
   - o Show emotional intelligence or have self-awareness.

     3. Machine learning features include:
   - ● **Learning from data:** ML models improve their performance by identifying patterns in data.
   - ● **Adaptability:** Models can adapt and improve over time with new data.

Touchpad Artificial Intelligence-XI (Answer Key) (Ver.3.0)   3

IMPRINT 1

- **Automation:** ML can automate decision-making processes without explicit programming.
- **Predictive capability:** ML models can predict future outcomes based on historical data.

4. Computer vision is a domain of AI that enables computers to interpret and understand visual information from the world, such as images and videos. It involves tasks like image classification, object detection, facial recognition, and scene reconstruction. A computer can 'see' images by converting them into numbers. Computer vision is used in applications such as self-driving cars, medical image analysis, and facial recognition systems.

5. Data refers to raw, unprocessed facts and figures without any context or meaning (e.g., numbers, text, images).

Information is data that has been processed, organized, or structured in a way that adds meaning and makes it useful for decision-making.

For example, "20°C" is data, but knowing that it is the current temperature in a specific city makes it information.

B. 1.
- **Supervised learning:** The model is trained on labelled data (input-output pairs). Example: Email classification as spam or not spam.
- **Unsupervised learning:** The model finds patterns from unlabelled data without specific outcomes. Example: Market segmentation to group customers.
- **Reinforcement learning:** The model learns by interacting with an environment, receiving rewards or penalties. Example: Game-playing AI, like AlphaGo.

Applications:
1. Fraud detection in banking
2. Self-driving cars
3. Product recommendations on e-commerce platforms
4. Predictive maintenance in manufacturing

2. Applications of Deep Learning are:
- **Automated Driving:** Deep learning is used to spot stoplights and traffic signals and also to detect pedestrians, reducing the incidence of accidents.
- **Aerospace and Defence:** Identifying objects from satellites and locate safe and unsafe zones for troops is another area where deep learning is playing a major role.
- **Medical Research:** Cancer researchers use deep learning to automatically detect cancer cells.
- **Industrial Automation:** Deep learning helps detect when people or objects are within an unsafe distance from the heavy machines thereby ensuring their safety.

3. a. **Snapchat image filters:** Belong to Computer Vision, which involves analysing and modifying images in real-time to apply effects like filters or facial alterations.
   b. **Virtual assistants like Siri and Alexa:** Belong to Natural Language Processing (NLP) and Speech Recognition, which allow the system to understand, interpret, and respond to voice commands.

4. ● Strong AI** refers to systems that have general intelligence, can understand and reason across a wide range of tasks, and possess consciousness. It remains theoretical and has not yet been achieved.

   ● Weak AI is designed for specific tasks and doesn't possess consciousness or understanding beyond those tasks. It focuses on narrow intelligence.

   Examples of Weak AI:

   1. Siri (virtual assistant)

   2. Google's search algorithms

5. Labelled data refers to a dataset where each data point is paired with the correct output (label). It is used in supervised learning, where the model learns to map inputs to the corresponding output. For example, in a labelled dataset of emails, each email is tagged as either "spam" or "not spam." The model uses these labels to learn the difference between spam and legitimate emails and applies this knowledge to classify new, unlabelled emails.

C. Competency-based/Application-based questions:

1. The type of AI learning described is Supervised Learning. In supervised learning, the algorithm is provided with labelled examples (such as images with correct classifications), and it learns to map inputs (e.g., images) to outputs (e.g., categories). Over time, the model generalizes from the examples to make accurate predictions on new, unseen data.

2. The learning type described for robots equipped with visual sensors is Reinforcement Learning. In reinforcement learning, an agent (in this case, a robot) learns from interacting with its environment. It takes actions based on the state of its surroundings and receives feedback in the form of rewards or penalties, which helps it improve its decision-making over time. Through trial and error, the robot learns optimal behaviours to achieve its goals.

Δi In Life (Page 136)

Do It Yourself

Δi Deep Thinking (Page 136)

Do It Yourself

 Δi Lab (Page 137)

Do It Yourself

Δi Ready 1

Do It Yourself

2. Unlocking Your Future in AI

Exercise

🔒 **Unsolved Questions**

SECTION A (Objective Type Questions)

Ai Quiz

A. 1. c   2. b   3. b   4. b   5. b   6. c

     7. c   8. a   9. b   10. a

B. 1. Numpy   2. communication   3. autonomous vehicles

     4. weapons   5. Datanami

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

SECTION B (Subjective Type Questions)

A. 1.
- Machine Learning Engineer
- Data Scientist
- AI Research Scientist
- Crop Yield Prediction Analyst
- Autonomous Vehicle Engineer

2. Python is popular for AI development due to its simplicity, readability, and extensive library support (e.g., TensorFlow, PyTorch, SciPy). It has a vast community, which provides excellent resources and support. Python's ability to handle complex computations, data structures, and machine learning tasks efficiently makes it an ideal choice for AI developers.

3. An NLP Specialist works on developing algorithms and systems that enable machines to understand, interpret, and generate human languages. Their work involves creating models for tasks such as language translation, sentiment analysis, and speech recognition. NLP is critical in applications like virtual assistants (e.g., Siri, Alexa) and chatbots.

4. KDnuggets is a leading platform for data science, artificial intelligence, machine learning, and big data. It provides valuable insights, tutorials, news, and trends in these fields, making it a key resource for professionals and researchers looking to stay updated on the latest advancements.

5.
- **Programming skills:** Proficiency in Python, R, or Java.
- **Machine Learning:** Understanding of algorithms and model training.
- **Shell scripting:** Adept at writing shell scripts to automate tasks. Enhances system efficiency and workflow management.
- **Mathematics and Statistics:** Knowledge of probability, linear algebra, and calculus.
- **Deep Learning:** Expertise in neural networks and frameworks like TensorFlow and PyTorch.

B. 1.   AI plays a transformative role in the financial sector by enhancing operations and decision-making. It is used for fraud detection, where machine learning algorithms analyse transaction patterns to identify suspicious activities. AI also powers automated trading systems, making real-time, data-driven investment decisions. In risk management, AI models assess creditworthiness and predict market trends. AI-driven chatbots improve customer service by handling inquiries efficiently. Overall, AI reduces operational costs, increases accuracy, and provides insights that enable more informed, faster decision-making.

2.   AI enhances customer experiences and optimizes operations in the retail industry. Personalized recommendations powered by AI analyse customer behaviour to suggest products, improving customer satisfaction and driving sales. AI is also used for demand forecasting, ensuring stock levels are optimized to meet consumer demand while reducing waste. Chatbots and virtual assistants provide 24/7 customer service, resolving issues quickly. Additionally, AI-driven automation in supply chains improves efficiency and cost-effectiveness. Overall, AI helps retailers deliver tailored experiences, streamline operations, and increase profitability.

3.   AI has significant applications in both government and military sectors, including surveillance, cybersecurity, and defence systems. In the military, AI powers autonomous drones and robotic systems, enhancing surveillance, reconnaissance, and target detection. It also assists in decision-making by analysing vast amounts of data. Governments use AI for smart city management, improving traffic flow and resource allocation. AI-driven cybersecurity systems protect sensitive data from cyber threats. However, these developments raise ethical concerns, such as privacy issues and the potential for autonomous weapons to escalate conflicts.

4.   In the AI-driven era, base skills such as problem-solving, critical thinking, and adaptability are crucial for career success. As AI continues to automate routine tasks, professionals need to focus on creative, strategic thinking that machines cannot replicate. Communication skills are essential for collaborating across teams and explaining complex AI concepts to non-technical stakeholders. Lifelong learning is important to stay updated on the latest AI advancements. These foundational skills ensure individuals can work alongside AI technologies, remain competitive, and contribute meaningfully to their fields.

5.   AI is reshaping agriculture by increasing efficiency and sustainability, contributing to global food security. Precision farming uses AI-driven sensors and drones to monitor crop health, soil conditions, and water levels, enabling optimized use of resources. AI-powered machines automate tasks such as planting, irrigation, and harvesting, reducing labour costs and improving yield. Predictive analytics help farmers anticipate weather patterns and crop diseases. By reducing waste and improving efficiency, AI has the potential to significantly enhance food production, ensuring sustainability and feeding a growing global population.

C.   **Competency-based/Application-based questions:**

1.   Akshat's specific responsibility as a Machine Learning Engineer is creating and training machine learning models for the AI application. His main focus is on developing models that can analyse customer browsing data and make accurate predictions about their preferences

to suggest relevant products. This aligns with his job of building data-driven models that help enhance customer experiences.

2. Trisha's main role as a Robotics Engineer is designing AI-powered robots that can perform tasks with precision and efficiency. In her case, she is working on a robotic arm for manufacturing, ensuring that it can carry out precise, repetitive tasks with high efficiency. Her work focuses on automation and optimization of industrial processes.

3. Shivani is demonstrating problem-solving skills in her approach to addressing the issue in her AI development project. By breaking down the problem into smaller parts, analysing them carefully, and then devising a comprehensive solution, she shows her ability to think logically and systematically resolve complex challenges.

4. Hemant's main responsibility as a Travel Recommendation Engine Developer is to use AI to offer personalized travel recommendations to customers. His job involves developing algorithms that analyse customer preferences and past travel history to suggest destinations, accommodations, and activities tailored to individual needs, improving customer satisfaction and engagement.

5. Vikas is demonstrating critical thinking in his analysis of the large dataset. By applying logical reasoning and critical analysis, he is able to accurately interpret the data, identify trends and patterns, and draw meaningful conclusions that contribute to the development of a new AI algorithm.

Δi In Life (Page 161)

Philosophy and AI are closely intertwined, as philosophy addresses many of the core questions and ethical dilemmas that arise in the development and application of artificial intelligence. Here are some key areas where philosophy plays a significant role in AI:

1. Ethics and Morality in AI

2. **Philosophy of Mind:** AI raises important questions about the nature of consciousness, intelligence, and mind.

3. **Free Will and Autonomy:** Should AI make decisions about human freedom or justice, and how do we limit AI's power?

4. Existential Risks and AI

5. **Epistemology and AI:** Philosophers question whether an AI that identifies patterns in data has true knowledge or if it's merely following statistical correlations.

Philosophy shapes the development, implementation, and regulation of AI by addressing fundamental questions about ethics, consciousness, autonomy, and existential risks. As AI continues to advance, the role of philosophy becomes even more critical in guiding its evolution in a way that aligns with human values and promotes responsible innovation.

Δi Deep Thinking (Page 161)

Do It Yourself

Do It Yourself

Ai Ready 2

Do It Yourself

3. Python Programming

Exercise

🔒 Unsolved Questions

SECTION A (Objective Type Questions)

Ai Quiz

A. 1. b   2. c   3. d   4. c   5. c   6. c

    7. b   8. a   9. c   10. b   11. a   12. d

    13. b   14. c   15. c   16. b

B. **1**. libraries, frameworks   2. Integrated Development and Learning Environment

    3. input()   4. +=   5. if-elif-else   6. immutable

    7. linear   8. comma   9. X = iris.data   10. IBM Skills Network

C. 1. False   2. True   3. True   4. True   5. True   6. False

    7. False   8. False   9. False   10. True

SECTION B (Subjective Type Questions)

A. 1. Operator precedence defines the order in which operators are evaluated in an expression. Higher-precedence operators are executed before lower-precedence ones. For example, multiplication '*' has higher precedence than addition '+', so in the expression 2 + 3 * 4, multiplication happens first, giving the result 14 instead of 20.

2.

Single Line Comment	Multiline Comment
Starts with a hash symbol (#)	Achieved by enclosing text in triple-quotes ('''or""")
Applicable to a single line	Usually spans multiple lines
Used for short, quick notes	Used for documentation, providing detailed explanations, or commenting out larger blocks of code

Single Line Comment	Multiline Comment

Example:
 	
`x=2 #single line comment`
	Example:
	
	```'''```
	`This is a multi-line comment.`
	`It can span multiple lines.`
	```'''```

3. KNN stands for K-Nearest Neighbors, a machine learning algorithm used for classification and regression. It works by finding the 'k' nearest data points to a given input and classifies or predicts based on the majority of the neighbours.

4. Dynamic typing means that the type of a variable is determined at runtime, not in advance. In Python, variables can change types during execution, which allows for flexibility and ease of writing code.

```
x = 2                    # x is an integer
x = "Hi!"                # Now x is a string
```

5. In a dictionary, keys are used to access values. Each key must be unique and immutable, while values can be mutable. Keys act like labels, helping to retrieve the corresponding value quickly.

   Example:

```
my_dict = {'name': 'Alka', 'age': 25}
print(my_dict['name'])  # Output: Alka
```

6. NumPy is useful in many real-life applications, such as:

   - **Scientific Computing:** Handling large datasets, simulations, and matrix operations.
   - **Image Processing:** Manipulating pixel data in image analysis.
   - **Machine Learning:** Efficiently handling data arrays and matrices for algorithms.
   - **Financial Analysis:** Performing statistical calculations on large sets of financial data.

7.
```
import numpy as np
    # From a list
    arr1 = np.array([1, 2, 3])
    # From a range of values
    arr2 = np.arange(0, 10, 2)
    # From zeros
    arr3 = np.zeros((2, 3))
    # From random values
    arr4 = np.random.rand(3, 3)
```

8. A DataFrame is a 2D data structure in pandas, similar to a table with rows and columns. One way to create it is from a dictionary:

```
import pandas as pd
data = {'Name': ['Anant', 'Billu'], 'Age': [25, 30]}
df = pd.DataFrame(data)
print(df)
```

9. The Iris Dataset is a well-known dataset used in machine learning and statistics. It contains 150 samples of iris flowers with four features (sepal length, sepal width, petal length, petal width) and a target label (species). It is commonly used to demonstrate classification algorithms.

10. ● **Feature:** Attributes or independent variables used as inputs to a model, such as sepal length in the Iris dataset.

   ● **Target value:** The output or dependent variable that the model is predicting, such as the species of the iris flower in the Iris dataset.

B. 1. Scikit-learn is a powerful Python library used for machine learning and data analysis. It offers a range of efficient tools for tasks like classification, regression, clustering, and dimensionality reduction. Some key features include:

   ● **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.

2. Splitting data into training and testing sets is essential to evaluate the performance of a machine learning model. The training set is used to train the model by learning the patterns in the data, while the testing set is kept aside to evaluate the model's generalization ability on unseen data. Without this separation, a model could overfit the training data, meaning it performs well on known data but poorly on new data. This split helps assess how well the model will perform in real-world scenarios.

3. Precision, confusion matrix, and accuracy are critical metrics used to evaluate the performance of classification models in machine learning.

   ● Precision measures the proportion of true positive predictions out of all positive predictions, showing how many predicted positives are actually correct.

- Confusion Matrix provides a detailed breakdown of true positives, true negatives, false positives, and false negatives, giving a comprehensive view of the model's performance.
- Accuracy measures the overall correctness of the model, calculating the percentage of correct predictions out of the total predictions.

These metrics are used together to assess how well the model handles both positive and negative classifications, especially when working with imbalanced datasets. For example, a high accuracy may not be sufficient if false positives are a concern, which is why precision and the confusion matrix help provide deeper insights.

4.
- **Adding Rows:** You can add rows using `pd.concat()` by concatenating a new DataFrame to the existing one.

  ```
  df = pd.concat([df, new_row])
  ```
- **Adding Columns:** You can assign a new column to a DataFrame:

  ```
  df['new_column'] = [values]
  ```
- **Deleting Rows:** Use the `drop()` function, specifying the index of the row to be dropped.

  ```
  df.drop(index=1, inplace=True)
  ```
- **Deleting Columns:** Similarly, columns can be dropped using `drop()`, specifying the axis.

  ```
  df.drop('column_name', axis=1, inplace=True)
  ```

5. Python offers a range of editors for coding, including:
- **IDLE:** Python's built-in editor, ideal for beginners.
- **PyCharm:** A popular Integrated Development Environment (IDE) with powerful features like debugging, testing, and intelligent code completion.
- **VS Code:** A lightweight, highly customizable code editor that supports Python and many extensions.
- **Jupyter Notebook:** A web-based environment primarily used for data analysis, machine learning, and visualization, allowing for interactive code execution.
- **Spyder:** An IDE geared towards data science and engineering, with features like variable explorer and a robust debugging system.

C. Competency-based/Application-based questions:

1. Benefits of using DataFrames over Excel:

   While Excel is widely used for data analysis, DataFrames (in Pandas) offer several advantages:
- **Efficiency with large datasets**: DataFrames can handle much larger datasets than Excel, offering better performance for computational tasks.
- **Automation**: DataFrames allow for easy automation of repetitive tasks through code, reducing the need for manual entry and lowering the risk of human error.

- **Advanced functionalities**: DataFrames support advanced data manipulation, such as filtering, joining, and grouping data, which can be done more efficiently than in Excel.
- **Integration with AI/ML**: DataFrames can easily integrate with machine learning libraries like Scikit-learn, making them ideal for data preprocessing and model training.
- **Reproducibility**: With DataFrames, you can write code that can be reused for other datasets, providing a more consistent and reproducible workflow than manual Excel processes.

2. The Iris dataset is an excellent choice for beginners in AI model testing for several reasons:
- **Well-known and simple:** The dataset is widely recognized and easy to understand, making it perfect for learning and experimentation.
- **Balanced and clean data:** It contains a balanced set of features (sepal length, sepal width, petal length, petal width) and well-labelled target classes (three species of iris flowers), reducing the need for extensive data cleaning.
- **Versatility for various models:** The Iris dataset works well with both classification and clustering algorithms, allowing students to explore different machine learning models and techniques.
- **Small size:** With just 150 samples, it's computationally efficient, meaning students can train and test models quickly without requiring heavy computational resources.

Using the Iris dataset would help grasp key AI concepts in a simpler and manageable way.

Δi In Life (Page 246)

Importance of the Iris Dataset in Real Life:

The Iris dataset is historically important because it was one of the first widely used datasets in the field of machine learning and pattern recognition. It remains a foundational dataset for teaching and experimentation for several reasons:

1. **Ease of use:** Its simplicity allows beginners to understand how machine learning algorithms work without being overwhelmed by complex data.
2. **Benchmarking:** Researchers use it to benchmark new algorithms and techniques, providing a standard comparison across studies.
3. **Insights into classification:** With its three classes of iris species and four easily distinguishable features (sepal length, sepal width, petal length, petal width), it's perfect for demonstrating classification algorithms.
4. **Introduction to supervised learning:** It offers a clear example of a classification problem where the output labels are well-defined, making it ideal for supervised learning demonstrations.

Drawbacks of the Iris Dataset:

1. **Limited complexity:** The Iris dataset is quite simple, with only 150 samples and 4 features, which limits its usefulness for advanced AI models or complex real-life scenarios.

2. **Overused:** Because of its popularity, the dataset has been extensively studied, meaning it's less challenging or useful for advanced machine learning practitioners looking for novel challenges.

3. **Small size:** With only 150 samples, the dataset may not represent real-world complexities, where datasets are much larger and more diverse.

5 Alternative Datasets to Iris:

1. **Wine Dataset:**
   - This dataset contains chemical analysis data of wines from three different cultivars. It includes 13 continuous features like alcohol content, color intensity, etc., and is often used for classification tasks.
   - **Available in:** Scikit-learn

2. **Breast Cancer Wisconsin Dataset:**
   - This dataset is used for binary classification of breast cancer (malignant vs. benign). It includes 30 numerical features based on cell characteristics.
   - **Available in:** Scikit-learn

3. **Digits Dataset:**
   - This dataset consists of 8x8 pixel images of handwritten digits, where the goal is to classify the digits (0-9). It's commonly used for multi-class classification problems.
   - **Available in:** Scikit-learn

4. **Titanic Dataset:**
   - A widely used dataset in machine learning, it contains information on passengers of the Titanic, including features such as age, fare, gender, and class, to predict survival rates.
   - **Available in:** Kaggle

5. **Boston Housing Dataset:**
   - This dataset includes various features related to houses in the Boston area, such as crime rate, number of rooms, and proximity to schools, making it suitable for regression tasks.
   - **Available in:** Kaggle

These datasets offer more complexity, real-world relevance, and diversity of features, making them ideal for moving beyond the basics of machine learning.

Δi Deep Thinking (Page 246)

AI Deep Thinking

	Numpy	Pandas
Data Compatibility	• Primarily designed to handle numerical data in the form of multi-dimensional arrays (ndarrays).	• Built on top of NumPy but provides support for heterogeneous data types, which means it can handle columns of

	Numpy	Pandas
	• Works well for homogeneous data types, where every element in the array is of the same data type (e.g., all integers or all floats). • It is ideal for mathematical and scientific computations that require fast array processing.	different types (e.g., integers, strings, and floats) in a single DataFrame. • Pandas is designed for data manipulation and can handle structured, tabular data (like spreadsheets and SQL tables), making it more versatile for data analysis tasks.
Performance	• Faster for numerical operations on large datasets due to its optimized implementation for array operations. • NumPy uses contiguous memory blocks, leading to more efficient memory use and faster operations, especially for tasks involving heavy mathematical computations.	• Pandas is slightly slower compared to NumPy when performing operations on pure numerical data because it adds more overhead due to its flexibility to handle heterogeneous data. • However, Pandas is optimized for tasks such as data manipulation, filtering, grouping, and merging of tabular data, which would be more cumbersome with NumPy.
Memory Usage	• More memory-efficient when dealing with large numerical datasets because it stores data in compact, continuous memory blocks. • Data types are typically homogeneous, so NumPy arrays are memory-dense, and operations on large datasets are generally faster.	• More memory-intensive than NumPy, as it supports multiple data types in a DataFrame (e.g., mixed integers, floats, and strings in different columns), leading to increased memory usage. • Pandas may also create copies of data during operations like indexing or filtering, which can consume more memory.
Ease of Use	Best suited for scientific computing and mathematical operations, but lacks intuitive ways to handle tabular data	Easier to use for data analysis, manipulation, and cleaning tasks. Its built-in methods for filtering, grouping, and merging data make it more suitable for handling real-world datasets.

- NumPy is faster and more memory-efficient for numerical operations and array-based computations, making it ideal for scientific and mathematical tasks.
- Pandas is more flexible and easier to use for handling structured, heterogeneous data with labelled axes, making it the go-to choice for data manipulation and analysis in real-world applications.

Δi Lab (Page 246)

Write code in Python for the following:

1.
```python
import numpy as np
zeros_array = np.zeros(10)
print(zeros_array)
```

2.
```python
import numpy as np          #need not be repeated if statement is
#already added
range_array = np.arange(20, 41)
print(range_array)
```

3.
```python
import pandas as pd
data = {'Name': ['Amit', 'Sangeeta', 'Manu', 'Priya'],
        'Age': [28, 32, 44, 24],
        'City': ['Delhi', 'Mumbai', 'Chennai', 'Kolkata']}
df = pd.DataFrame(data)
print(df)
```

4.
```python
import pandas as pd
data = {'Name': ['Amit', 'Esha', 'Ravi', 'Priya'],
        'Age': [25, 30, 28, 24],
        'City': ['Delhi', 'Mumbai', 'Chennai', 'Kolkata']}
df = pd.DataFrame(data)
print(df[['Name', 'City']])
```

5.
```python
import pandas as pd
data = {'Name': ['Amit', 'Esha', 'Ravi', 'Priya'],
        'Age': [25, 30, 28, 24],
        'City': ['Delhi', 'Mumbai', 'Chennai', 'Kolkata']}
df = pd.DataFrame(data)
```

```
        df['Marks'] = [68, 40, 78, 82]
        print(df)
    6.  import pandas as pd
        data = {'Name': ['Amit', 'Esha', 'Ravi', 'Priya'],
                'Age': [25, 30, 28, 24],
                'City': ['Delhi', 'Mumbai', 'Chennai', 'Kolkata']}
        df = pd.DataFrame(data)
        df.drop(columns=['Age'], inplace=True) #drop works on original
            #dataframe
        print(df)
    7.  import pandas as pd

        # Create a DataFrame
        data = {'City': ['Mumbai', 'Delhi', 'Bangalore', 'Hyderabad',
        'Ahmedabad'],
               'Population': [12442373, 10927986, 8436675, 6809970, 5570585],
                'State': ['Maharashtra', 'Delhi', 'Karnataka', 'Telangana',
                'Gujarat']}

        df = pd.DataFrame(data)

        # Display the DataFrame
        print(df)

        # Display the data types of columns
        print(df.dtypes)
    8.  from sklearn.datasets import load_iris
        import pandas as pd

        # Load Iris Dataset
        iris = load_iris()
```

```
    df = pd.DataFrame(iris.data, columns=iris.feature_names)
    df['species'] = iris.target


    # Print the first 5 rows of the dataset
    print("First 5 rows of the dataset:")
    print(df.head())


    # Print the names of the features
    print("\nNames of the features (attributes) in the dataset:")
    print(iris.feature_names)


    # Print the target variable names (species)
    print("\nTarget variable names (species) in the dataset:")
    print(iris.target_names)
```

9.
```
   from sklearn.datasets import load_iris
   import pandas as pd


   # Load Iris Dataset
   iris = load_iris()
   df = pd.DataFrame(iris.data, columns=iris.feature_names)
   df['species'] = iris.target


   # Print summary statistics of the dataset
   print("Summary statistics of the dataset:")
   print(df.describe())
```

10. a. 7.5
     b. 8
     c. 3   3

Δi Ready 3

Do It Yourself

4. Introduction to Capstone Project

Exercise

🔒 **Unsolved Questions**

SECTION A (Objective Type Questions)

Ai Quiz

A. 1. c   2. c   3. b   4. c   5. c   6. c

     7. b   8. b

B. 1. opportunities   2. educated individual   3. 40

     4. population, dependence   5. The ocean   6. Problem definition

     7. adjustment, refinement   8. Data exploration

C. 1. True   2. False   3. False   4. True   5. True   6. False

     7. True   8. True

SECTION B (Subjective Type Questions)

A. 1. The second stage of Design Thinking is the Define stage. In this phase, designers integrate the information gathered during the Empathize stage to clearly define the problem they are solving. This involves creating a problem statement that is human-centred and reflects the needs of the users. The aim is to express the challenge in a way that guides the team toward innovative and user-driven solutions.

2. Evaluation is the stage in an AI project where the performance of the model is assessed using metrics like accuracy, precision, recall, and F1-score to ensure the model meets desired objectives. In contrast, Deployment is the stage where the trained AI model is integrated into real-world applications or production environments to be used by end-users. Evaluation ensures the model's quality, while deployment makes the model accessible for use.

3. SDG 14 focuses on conserving and sustainably using the oceans, seas, and marine resources, which are critical for supporting billions of livelihoods and maintaining biodiversity. SDG 15 emphasizes protecting, restoring, and promoting sustainable use of terrestrial ecosystems, combating deforestation, and halting biodiversity loss. Both goals are essential for environmental sustainability and ensuring the health of ecosystems that are crucial for human survival and economic stability.

4. AI modelling involves the process of building, training, and deploying machine learning or deep learning models to solve specific tasks. These models use data to learn patterns, make predictions, and automate decision-making processes. AI modelling typically includes data collection, model selection, training, and evaluation phases to create an accurate system for tasks like image recognition, language translation, or recommendation systems.

5. A capstone project is a final project that students undertake at the end of an academic program or course. It integrates the knowledge and skills they have learned throughout their studies, allowing them to solve a real-world problem or explore a complex topic. In data science or AI, a capstone project often involves analysing data, developing models, and presenting findings to demonstrate the practical application of the skills acquired during the program.

B. 1. Deployment in the context of AI and machine learning refers to the process of making a trained model available for use in a real-world environment. After the model has been developed, trained, and evaluated, it is integrated into a system or application where it can process new data and generate predictions or decisions. Deployment involves configuring the model for scalability, ensuring it interacts correctly with other systems, and monitoring its performance to ensure reliability. It is a critical step to ensure the AI solution is functional and accessible to end-users.

2. Evaluation of an AI model involves assessing its performance on unseen data by using metrics such as accuracy, precision, recall, and F1-score. This process is crucial because it helps determine how well the model generalizes to new data, ensuring that it is reliable and effective for practical use. Without proper evaluation, a model may produce inaccurate or biased results, which could lead to flawed decision-making. Evaluation is important to detect overfitting, identify errors, and optimize the model for better performance in real-world scenarios.

3. Brainstorming is popular with design teams because it encourages the free flow of ideas in a collaborative and non-judgmental environment. It helps generate a wide range of creative solutions to a problem, fostering innovation by allowing team members to build on each other's ideas. Brainstorming also breaks down barriers between different perspectives, which can lead to more comprehensive and unique solutions. The process is particularly valuable in the early stages of design thinking, where diverse ideas are needed to address complex challenges.

4. The 5W and 1H method is a systematic questioning technique used to gather comprehensive information about a situation or problem. It involves asking six key questions: Who, What, When, Where, Why, and How. These questions help break down the problem by focusing on different aspects of it, ensuring that all relevant factors are considered. For example, "Who is involved?" "What is the issue?" "When does it occur?" "Where does it happen?" "Why is it important?" and "How can it be resolved?" This method is effective in identifying root causes and developing actionable solutions.

5. Yes, identifying a problem is often more important than providing a solution because a well-defined problem leads to more effective and targeted solutions. If the problem is not understood correctly, even the best solutions might address the wrong issue. For example, in business, a drop in sales might be due to poor customer service, not product quality. Without identifying the root cause (the problem), efforts to improve the product might not help.

Albert Einstein once said, "If I had an hour to solve a problem, I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions," emphasizing the importance of problem identification.

C. **Competency-based/Application-based questions:**

1. In response to Anuj's query, I would explain that the Ideate and Prototype stages of Design Thinking require the most resources and effort to ensure that a practically implementable prototype is developed. Reasons are as follows:

   - **Ideate Stage:** This stage involves brainstorming and generating a wide range of creative ideas and potential solutions. It's crucial to explore as many possibilities as possible so that you can narrow down the best ones. The more time and resources spent here, the better the quality of the ideas.

   - **Prototype Stage:** After generating ideas, creating prototypes (early, scaled-down versions of the solutions) helps to visualize and test those ideas. Building a good prototype requires resources to ensure its functional enough for testing, while still being cost-effective. The goal is to create something tangible that can be tested and improved.

   By investing effort in these stages, Anuj will be able to minimize time during the actual project execution because he will already have a clear, tested, and validated solution that meets users' needs.

2. In order to build an effective AI system to predict traffic patterns, the data scientists must carefully consider and collect various key factors that influence traffic. Here are the essential factors to ascertain:

   - **Time of Day:** Traffic patterns vary significantly depending on the time of day (e.g., rush hours, late-night hours), making it a crucial variable.

   - **Day of the Week:** Weekdays and weekends typically have different traffic volumes, so understanding this distinction is vital.

   - **Weather Conditions:** Factors like rain, snow, fog, or clear skies can heavily impact traffic flow and must be factored into the model.

   - **Road Conditions:** Data on road quality, construction work, accidents, or closures is critical for accurate predictions.

   - **Traffic Volume and Flow:** Historical data on the number of vehicles on the road, traffic jams, and flow rates provide key insights into patterns.

   - **Special Events:** Events like rain, concerts, sports games, or festivals can drastically change normal traffic behaviour in specific areas.

   - **Public Transportation Data:** Information about bus, metro, or train schedules can affect road traffic, especially during peak hours.

   - **Geographic Factors:** The layout of roads, intersections, bottlenecks, and alternate routes in different parts of the city can influence traffic patterns.

- **Accident Reports:** Historical data on accidents, their locations, and impact on traffic must be included to anticipate disruptions.
- **Traffic Control Measures:** Signals, traffic lights, speed limits, and traffic management systems also play a role in managing traffic flow.

By collecting accurate data on these factors and properly defining them as features, the AI system can make more reliable and informed traffic predictions for drivers and city planners.

Δi In Life (Page 273)

Some Sustainable Development Goals (SDGs) are universally important for both developed and developing nations due to their global relevance and impact. These SDGs include:

1. **SDG 3:** Good Health and Well-being – Ensuring healthy lives and promoting well-being for all at all ages.
2. **SDG 4:** Quality Education – Ensuring inclusive and equitable quality education and promoting lifelong learning opportunities.
3. **SDG 7:** Affordable and Clean Energy – Ensuring access to affordable, reliable, sustainable, and modern energy for all.
4. **SDG 8:** Decent Work and Economic Growth – Promoting sustained, inclusive, and sustainable economic growth, full and productive employment, and decent work for all.
5. **SDG 9:** Industry, Innovation, and Infrastructure – Building resilient infrastructure, promoting inclusive and sustainable industrialization, and fostering innovation.
6. **SDG 13:** Climate Action – Taking urgent action to combat climate change and its impacts.
7. **SDG 12:** Responsible Consumption and Production – Ensuring sustainable consumption and production patterns.

Waste Tyres Linked to SDGs:

Waste tyres pose both environmental and health risks but can also be linked to several SDGs if managed sustainably:

i. **SDG 11:** Sustainable Cities and Communities – Proper management of waste tyres contributes to making cities and human settlements inclusive, safe, resilient, and sustainable by reducing landfill and pollution.
ii. **SDG 12:** Responsible Consumption and Production – Recycling and reusing waste tyres help in promoting sustainable consumption and reducing waste.
iii. **SDG 13:** Climate Action – Waste tyre recycling reduces the need for producing new materials, lowering greenhouse gas emissions and mitigating climate change impacts.
iv. **SDG 9:** Industry, Innovation, and Infrastructure – Innovation in recycling processes and creating new products from waste tyres supports sustainable industrialization and fosters innovation.
v. **SDG 15:** Life on Land – Preventing the dumping of tyres in the environment helps protect ecosystems, reducing pollution and damage to terrestrial life.

By addressing waste tyre management, countries can contribute to multiple SDGs and promote sustainable development.

Δi Deep Thinking (Page 374)

1. By incorporating sustainable development principles into capstone projects across various academic disciplines, students can directly address global challenges such as climate change, poverty, and inequality. These projects can propose innovative solutions that align with the UN SDGs, such as promoting clean energy (SDG 7), responsible consumption (SDG 12), and poverty reduction (SDG 1). When students apply sustainability in real-world scenarios, they help advance these goals while fostering awareness and actionable outcomes in areas like environmental conservation, social justice, and economic growth, contributing to a more sustainable future.

2. A capstone project that integrates knowledge from multiple disciplines, such as engineering, business, and social sciences, offers practical solutions to real-world problems. This multidisciplinary approach enhances students' problem-solving abilities, critical thinking, and teamwork skills, preparing them for professional environments where such integration is key. By solving real-world challenges, students gain experience that is highly relevant to employers, making them more competitive in the job market. Additionally, these projects often address societal needs, showcasing how academic learning can contribute to broader community and global development goals.

Δi Lab (Page 374)

Do It Yourself

Δi Ready 4

Do It Yourself

5. Data Literacy—Data Collection to Data Analysis

Exercise

🔒 Unsolved Questions

SECTION A (Objective Type Questions)

Δi Quiz

A.	1. b	2. b	3. b	4. c	5. c	6. b
	7. b	8. b	9. b	10. c	11. b	12. b
	13. a	14. d				

B. 1. decision-making   2. Questionnaire   3. Data analysts

4. Median   5. line graph   6. Binning

7. training   8. Cross-validation

C. 1. True   2. False   3. True   4. True   5. False   6. True

7. False   8. False

SECTION B (Subjective Type Questions)

A. 1. Testing data is a subset of the original dataset used to evaluate the performance of a machine learning model after it has been trained. It consists of unseen data that the model did not use during training, allowing for an unbiased assessment of its accuracy and generalization ability. The results from testing data help determine how well the model performs in real-world scenarios.

2. Ordinal data represents variables with ordered categories, where the order matters but the intervals between the categories are not equal or meaningful. For example, rankings like "first", "second", and "third" in a race are ordinal, as they indicate relative positions, but the difference in performance between ranks is not necessarily consistent or measurable.

3. A pie chart is a circular graph used to represent data as slices of a pie, where each slice corresponds to a category's proportion within the whole. For example, if you survey students about their favourite fruit and 50% choose apples, 30% choose bananas, and 20% choose oranges, a pie chart will visually show these proportions with different-sized slices for each fruit.

4. To install the Matplotlib library in Python, the following command is used:

pip install matplotlib

5. A matrix is a rectangular array of numbers, symbols, or expressions arranged in rows and columns. It is widely used in mathematics and computer science for various applications, including solving systems of linear equations, performing transformations, and representing data in machine learning models. For example, a 2x2 matrix looks like:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} => \begin{bmatrix} 3 & -2 \\ 4 & 8 \end{bmatrix}$$

B. 1. Secondary data refers to data collected by someone else and made available for others to use. Key sources include:

- **Social Media Data Tracking:** Collecting information from social media sites such as user postings, comments, and interactions.

- **Web scraping:** Employing automated software to gather specific information and data from websites.

- **Satellite Data Tracking:** Collecting data on the Earth's surface and atmosphere through satellite.

- **Online data platforms:** Collecting data from websites that provide pre-compiled datasets for diverse purposes.

- **Books, Textbooks and Encyclopaedias:** Information that has been researched, compiled, and written by authors or editors.

These sources offer valuable information for research, but may not always be tailored to specific needs.

2. - **Structured Data:** Highly organized data that follows a fixed schema, typically found in databases. Example: Excel spreadsheets or SQL tables with defined rows and columns.
   - **Semi-structured Data:** Data with some organizational properties, but lacking a rigid schema. Example: JSON or XML files, which have tags but allow variable data organization.
   - **Unstructured Data:** Data that has no predefined structure or organization. Example: Text documents, images, videos, or social media posts.

Each type requires different processing methods for analysis in machine learning or data science.

3. Data transformation is the process of converting raw data into a suitable format 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.

4. - **Neural Networks:** Matrices represent weights and activations in neural networks, enabling calculations of inputs and outputs through matrix multiplications.
   - **Image Processing:** Images are stored as pixel matrices, and AI models use matrix operations for tasks like filtering, transformations, and edge detection.
   - **Natural Language Processing (NLP):** Word embeddings are represented as vectors, with matrix multiplication used to transform them during language model training.
   - **Sales Forecasting and Price Prediction:** Matrices are used to represent relevant predictor and response variables.

5.

Training Data	Testing Data
Training data is used to build and train machine learning models.	Testing data is used to evaluate the model's performance.
Training data helps the model learn patterns and features.	Testing data assesses how well the model generalizes to unseen data.
Training data is exposed to the model during the learning phase.	Testing data remains unseen until the evaluation phase.
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.

C. Competency-based/Application-based questions:

1. To visualize the data of maximum temperatures across months, a line graph would be a suitable method, as it effectively represents changes over time and helps to observe trends.

```
import matplotlib.pyplot as plt

# Data for the months and maximum temperatures
months = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug',
'Sep', 'Oct', 'Nov', 'Dec']
max_temperatures = [19.50, 23.00, 27.95, 34.97, 37.66, 36.26, 32.90,
31.77, 31.73, 30.92, 27.33, 22.77]

# Plotting the line graph
plt.figure(figsize=(10, 6))
plt.plot(months, max_temperatures, marker='o', linestyle='-',
color='b')

# Adding labels and title
plt.title('Maximum Temperature by Month', fontsize=14)
plt.xlabel('Month', fontsize=12)
plt.ylabel('Max. Temperature (°C)', fontsize=12)

# Show the plot
plt.grid(True)
plt.show()
```
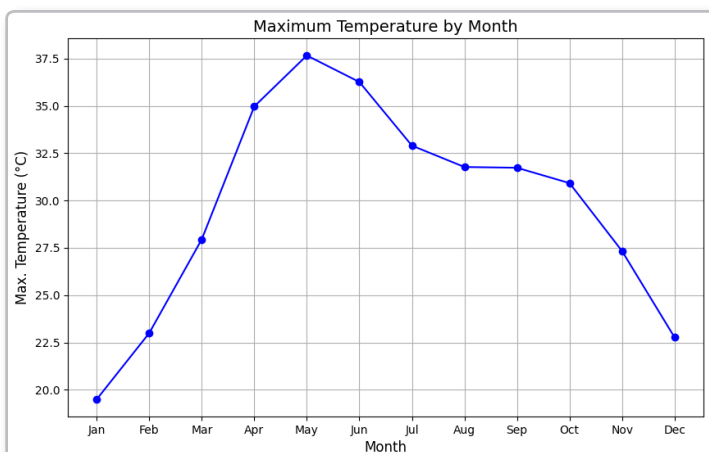
2. The variance and standard deviation for both classes are as follows:
   - Class A:
     - o Variance: 6.16
     - o Standard Deviation: 2.48
   - Class B:
     - o Variance: 6.00
     - o Standard Deviation: 2.45

   Both classes have very similar variance and standard deviation, but Class B has a slightly lower variance and standard deviation, indicating that Class B's scores are marginally more consistent than Class A's.

Δi In Life (Page 316)

Here are five more important statistics concepts students should know for a career in Data Science:

1. **Probability Distributions:** Understanding various probability distributions like Normal, Binomial, and Poisson distributions is crucial for modelling and predicting data patterns.
2. **Hypothesis Testing:** This involves testing an assumption about a population parameter. Common tests include t-tests, chi-square tests, and ANOVA, which help make inferences from data.
3. **P-value:** The p-value helps to determine the significance of your results in hypothesis testing, showing whether the observed result occurred by chance.
4. **Regression Analysis:** Techniques like linear and logistic regression are essential for predicting relationships between variables and making forecasts.
5. **Confusion Matrix and Precision/Recall:** These concepts are key for evaluating the performance of classification models, especially in measuring how well a model distinguishes between different classes.

Δi Deep Thinking (Page 316)

AI Deep Thinking

Here are five key characteristics of quality data that are essential for fostering a culture of high-quality data within an organization:

1. **Accuracy:** Data should be correct and free from errors or distortions. Accurate data ensures that insights and decisions based on it are reliable and reflect the true state of the business or environment.
2. **Consistency:** Data should be uniform across different systems and datasets. Consistent data allows for seamless integration, comparison, and analysis without discrepancies between sources.
3. **Completeness:** All required data should be present and available. Missing or incomplete data can lead to gaps in analysis and inaccurate conclusions.

4. **Timeliness:** Data should be up-to-date and available when needed. Timely data ensures that decisions are made based on the most current information, reducing the risk of outdated insights.

5. **Relevance:** The data collected should be applicable and useful for the specific analysis or decision-making task at hand. Irrelevant data adds noise and can detract from actionable insights.

Δi Lab (Page 316)

Do It Yourself

Δi Ready 5

Do It Yourself

6. Machine Learning Algorithms

Exercise

🔒 Unsolved Questions

SECTION A (Objective Type Questions)

Δi Quiz

A. 1. d

2. d (**Note:** Siri and Alexa are quite advanced. Supervised learning helps them understand and interpret user commands by training on labelled data, while reinforcement learning enables them to improve responses over time based on user interactions and feedback. Additionally, unsupervised learning may be used for tasks like identifying patterns in user behaviour to make personalized suggestions.)

    3. b   4. b   5. a   6. d   7. c   8. c

    9. c   10. c

B. 1. prediction   2. dependent   3. distinct   4. Causation

    5. centroid   6. clusters   7. slope   8. -1

    9. Correlation   10. continuous

C. 1. False   2. True   3. False   4. False   5. False   6. False

    7. True   8. True   9. True   10. False

SECTION B (Subjective Type Questions)

A. 1. 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.

2. In **supervised learning**, models are trained using labelled data where the input-output pairs are known, allowing the system to learn the mapping function. In **unsupervised learning**, the model works with unlabelled data, seeking to identify hidden patterns or groupings without predefined outputs.

3. 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.

4. The line of least squares in linear regression is the best-fitting line that minimizes the sum of the squared differences (errors) between the observed data points and the predicted values. It represents the relationship between the independent variable (x) and the dependent variable (y). The goal of the least squares method is to minimize these errors, ensuring the line is as close to the data points as possible. This line is used to make predictions by projecting the relationship between variables based on historical data. It is expressed by the equation $(y = mx + b)$, where 'm' is the slope and 'b' is the intercept.

5. Advantages of Linear Regression
- Linear regression is a simple technique and easy to implement.
- Efficient to train the machine on this model.

Disadvantages of Linear Regression
- 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).

6. "Correlation is not Causation" means that while two variables may show a statistical relationship (correlation), this does not imply that one variable directly causes the other to change. For example, we might find a correlation between coffee consumption and productivity levels, but this doesn't mean drinking more coffee directly causes increased productivity. There could be other factors, such as work environment or sleep quality, influencing both. Regression can highlight relationships between variables, but it doesn't prove cause-and-effect.

7. Interpretation of Correlation Coefficient:
- **r = 1:** Perfect positive correlation, indicating that as one variable increases, the other also increases proportionally.

- **r = -1:** Perfect negative correlation, meaning as one variable increases, the other decreases proportionally.
- **r = 0:** No correlation, implying no linear relationship between the variables.

8. **Classification:** Labeled data is used, and points are divided into predefined classes (e.g., separating cats and dogs).

   Clustering: Unlabeled data is grouped into clusters based on similarity (e.g., grouping animals by features).
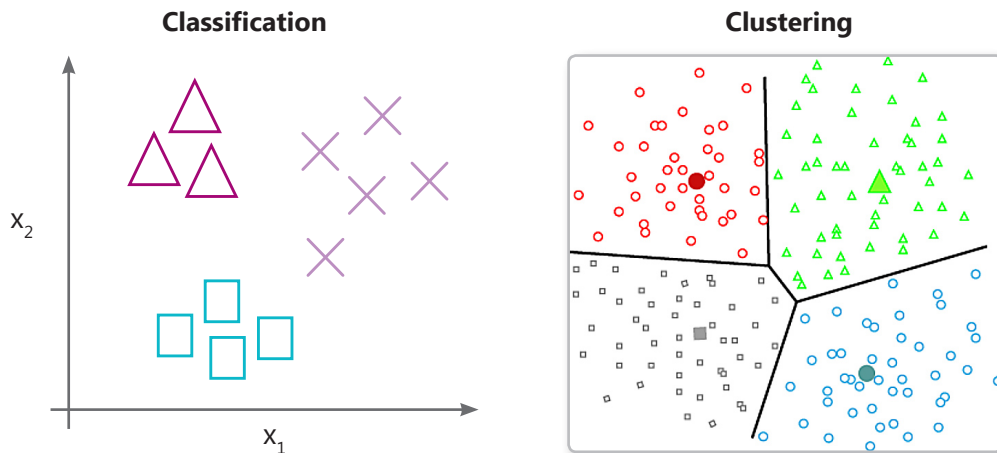
Graphically, classification shows distinct class boundaries, while clustering displays naturally forming groups without predefined labels.

B. 1. The classification process in machine learning involves predicting the category or class label of new data based on patterns learned from labelled training data. It begins with gathering and preprocessing the data to remove noise and normalize features. Next, a classification algorithm, such as decision trees, K-nearest neighbours (KNN), or support vector machines (SVM), is selected. The model is trained using labelled data where the correct class is already known. During training, the algorithm learns to identify patterns and relationships. Once trained, the model can classify new, unseen data by assigning it to one of the predefined categories.

2. Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship, where changes in the independent variable(s) predict changes in the dependent variable. The goal is to fit a straight line ($y = mx + b$) that best represents the data.

   Applications of linear regression:
   - Predicting house prices based on features like size and location
   - Forecasting sales revenue
   - Analyzing the relationship between variables in finance
   - Estimating medical outcomes based on patient data.

3. The four assumptions are:
   - For linear regression, the two variables should be continuous, such as height, temperature, or income.
   - A linear relationship between the variables is essential, which can be checked using a scatterplot.
   - Outliers, or data points that deviate significantly from others, can distort results and should be minimized or removed, as they affect the line of best fit and correlation.
   - Additionally, the variables should be approximately normally distributed for accurate analysis. Outliers and irregularities can lead to unreliable conclusions, so handling them carefully is crucial.

4. Binary classification involves classifying data into two distinct categories or classes, such as "yes" or "no," "positive" or "negative." For example, a spam filter classifies emails as either "spam" or "not spam."

   Multiclass classification, on the other hand, deals with problems where there are more than two classes. For example, classifying images into categories like "cat," "dog," or "bird." In binary classification, there are only two outcomes, whereas in multiclass, the model must handle multiple possible outcomes.

5. The steps of the algorithm are as follows:

   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.

C. Competency-based/Application-based questions:

1. KNN is a supervised learning algorithm used for classification, where the label of new data is predicted based on its neighbours. For example, predicting if an email is spam. K-means is an unsupervised learning algorithm for clustering, grouping data into clusters without labels. For example, grouping customers based on purchasing behaviour.

2. The SmartCommute system uses reinforcement learning, a type of machine learning where the model learns and adapts to optimize decisions, in this case, adjusting bus routes based on real-time feedback like traffic conditions.

3. c. A is true but R is false.

4. b. Both A and R are true but R is not the correct explanation of A.

5. d. Statement 1 is FALSE, Statement 2 is TRUE

6. a. Statement 1 is TRUE, Statement 2 is TRUE

Δi In Life (Page 363)

Do It Yourself

Δi Deep Thinking (Page 364)

Students can research in the following areas:

- Risk Analysis
- Prediction of Shelf Life
- Quality Control
- Optimized Decision Making
- Cost Management

Δi Lab (Page 364)

Do It Yourself

Δi Ready 6

Do It Yourself

7. Leveraging Linguistics and Computer Science

Exercise

🔒 Unsolved Questions

SECTION A (Objective Type Questions)

Δi Quiz

A.	1. c	2. c	3. c	4. a	5. a	6. c
	7. c	8. c	9. a	10. c	11. b	12. b

B. 1. closely   2. NLP   3. Sentiment analysis, emotion detection

4. Chatbots   5. parts of speech   6. Emotion detection

7. intents, entities, responses   8. certainty   9. frontend

10. phrase/statement

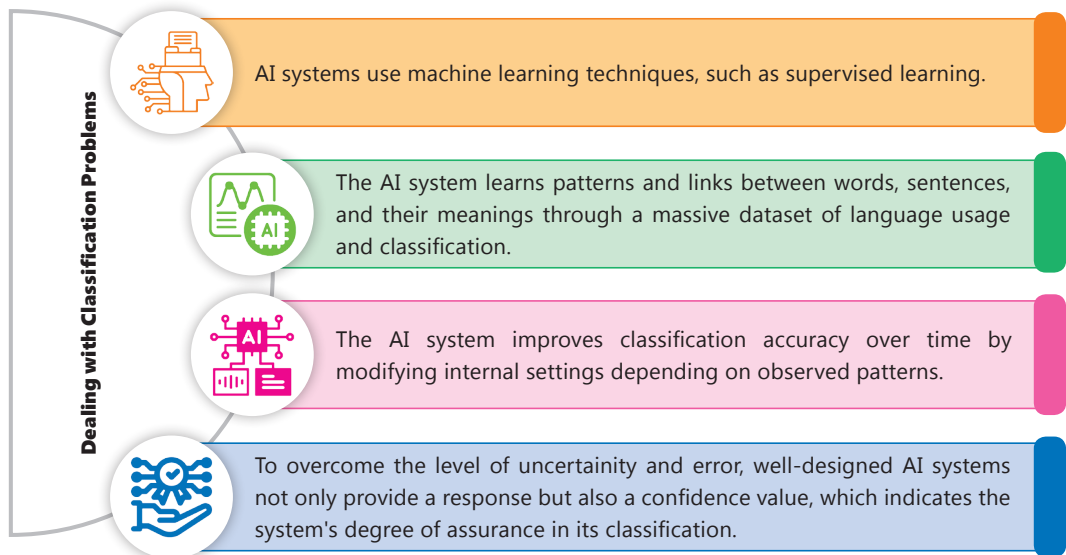| C.	1. False	2. False	3. True	4. True	5. True	6. False
	7. False	8. False	9. True	10. True		

SECTION B (Subjective Type Questions)

A. 1. Coreference resolution is the process in NLP where the system determines when different expressions refer to the same entity. For example, in the sentence, "Suman ate a pastry. She enjoyed it," "She" refers to "Suman," and "it" refers to the "pastry." Coreference resolution helps the system understand these references and ensure correct interpretation.

2. (a) **Intent:** In the context of NLP, intent is the purpose or goal behind a user's input, representing what the user wants to achieve, like booking a flight or asking for weather updates.

   (b) **Tokens:** Tokens are individual units of a text, such as words, numbers, or punctuation marks, used in NLP to break down a sentence into smaller parts for analysis.

3. **Sentiment Analysis:** It is used to determine the emotional tone behind a body of text, often used in social media monitoring or customer feedback analysis.

   Voice Assistants: We use voice assistants for making calls, setting alarms, scheduling meetings, making notes, and browsing the internet. To comprehend and react to human speech, these voice assistants employ a complex fusion of speech recognition, natural language processing, and natural language interpretation.

4. The statement refers to the fact that the meaning of words often depends on context. For example, the word "bank" can mean a financial institution or the side of a river. Without context, the intended meaning can be unclear or ambiguous.

5. The five phases of NLP are:
   - Lexical Analysis
   - Syntactic Analysis (Parsing)
   - Semantic Analysis
   - Discourse Integration
   - Pragmatic Analysis

B. 1. Document analysis is an important application of NLP, used to process, classify, and extract relevant information from large volumes of text. It helps in organizing and categorizing documents based on keywords, topics, or themes. This technology is applied in areas like risk management, legal document review, and automated contract analysis. Document analysis also supports summarization, making it easier for businesses to sift through lengthy documents and retrieve essential information efficiently.

2. The stages of NLP are as follows:
   - **Lexical Analysis:** Breaking down text into tokens such as words or phrases.
   - **Syntactic Analysis (Parsing):** Analysing sentence structure to understand grammar and word order.
   - **Semantic Analysis:** Assigning meaning to each word based on the context.

- **Discourse Integration:** Understanding how individual sentences are connected to form a coherent narrative.
- **Pragmatic Analysis:** Interpreting the meaning of text in relation to external knowledge and context.

3. 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.
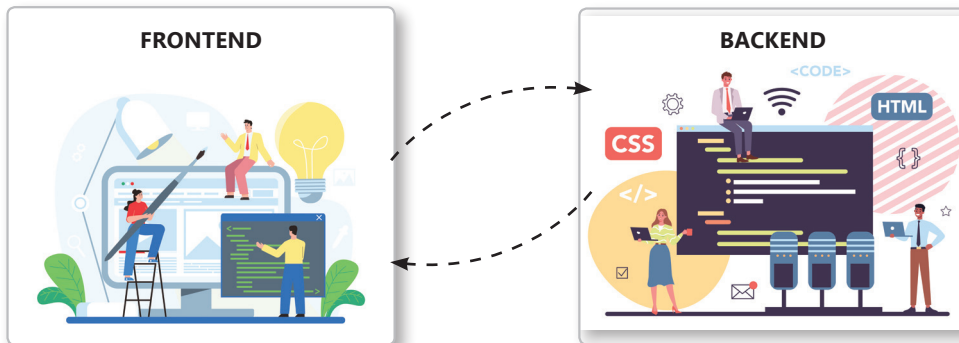
4. In chatbots, a dialog refers to the back-and-forth interaction between a user and the system. For example, in an e-commerce chatbot, a user might ask, "Do you have blue jeans?" The chatbot responds, "Yes, we have blue jeans in stock. What size are you looking for?" This sequence of conversation is the dialog, structured around intents and responses.

5.

Dealing with Classification Problems

AI systems use machine learning techniques, such as supervised learning.

The AI system learns patterns and links between words, sentences, and their meanings through a massive dataset of language usage and classification.

The AI system improves classification accuracy over time by modifying internal settings depending on observed patterns.

To overcome the level of uncertainty and error, well-designed AI systems not only provide a response but also a confidence value, which indicates the system's degree of assurance in its classification.

C.   **Competency-based/Application-based questions:**

1. Challenges in the email classification problem:

   - Dealing with ambiguous content. For example, an AI system might struggle to distinguish between a legitimate promotional email and a spam message, as both can contain similar phrases like "exclusive offer" or "limited time deal."
   - Emails that use clever language to bypass filters, such as deliberately misspelled words.
   - The evolving nature of spam, requiring constant retraining of the AI to recognize new patterns. For example, a message with "You won a prize!" may initially appear legitimate, but over time, this becomes associated with spam.

2. Transitioning from a manual support system to a chatbot, whether rule-based or AI-based, involves challenges. Rule-based chatbots can handle simple queries but may struggle with complex issues, leading to frustration. AI-based chatbots are more flexible but require substantial training data and may misinterpret nuanced language. Staff must adapt to managing the chatbot and handling escalations.

   Despite these hurdles, chatbots significantly improve efficiency by reducing response time and handling repetitive tasks, allowing the team to focus on more complex inquiries. This ultimately enhances customer satisfaction through faster issue resolution, though initial adjustments are necessary.

3. a. Both A and R are true and R is the correct explanation of A.

Δi Deep Thinking (Page 397)

Students can research on the following challenges/concerns related to voice assistants:

- Privacy Issues
- Security Risks
- Reliability
- Accessibility Issues
- Biases in Responses
- Potential Misuse
- Data Ownership and Consent
- Dependence and Over-reliance

Δi Lab (Page 397)

Do It Yourself

Δi Ready 7

Do It Yourself

8. AI Ethics and Values

Exercise

🔒 **Unsolved Questions**

SECTION A (Objective Type Questions)

Δ**ⅈ** **Q**uiz

A. 1. b   2. b   3. c   4. b   5. c   6. c

     7. c   8. d   9. b   10. c

B. 1. weather   2. visible   3. Fairness   4. Transparency

     5. Accountability   6. Privacy   7. Robustness   8. training

C. 1. False   2. False   3. True   4. False   5. True   6. True

SECTION B (Subjective Type Questions)

A. 1. 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. 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.

2. Several measures can be taken to ensure fair treatment by AI systems:

   - **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.

3. The main goals of AI ethics are to ensure responsible and fair AI development, minimizing bias, discrimination, and privacy violations while maximizing benefits like decision-making and inclusivity. It focuses on fairness, transparency, and accountability, ensuring AI treats individuals equitably and operates safely within society.

4. An example of AI bias in healthcare is in predictive algorithms that prioritize patient care. If trained on historical data reflecting systemic inequalities, these systems may allocate fewer resources to minority patients, propagating healthcare inconsistencies. This bias leads to unequal access to timely, quality care and reinforces existing inequalities in the healthcare system.

5. Bias awareness in AI technologies is crucial because biased AI systems can lead to unfair and discriminatory outcomes, reinforcing existing inequalities. AI systems often learn

from historical or unrepresentative data, which may contain biases. Without recognizing and addressing these biases, AI can disproportionately impact certain groups, leading to negative consequences in areas like hiring, healthcare, and criminal justice. Ensuring fairness and accountability in AI helps promote ethical, inclusive, and responsible technology use in society.

B. 1. Cognitive bias: refers to systematic errors in thinking that affect decision-making, often leading people to make judgments based on subjective experiences rather than objective facts. It can influence decisions by causing individuals to rely on mental shortcuts or preconceived notions, resulting in skewed perceptions and irrational conclusions. In AI, cognitive biases embedded in data or models can perpetuate flawed outcomes, affecting fairness and accuracy.

2. Games like the Moral Machine and Survival of the Best Fit help individuals understand ethical issues and biases in AI by simulating real-world dilemmas. The Moral Machine allows players to make moral decisions involving autonomous vehicles, exposing them to ethical conflicts and societal expectations. Survival of the Best Fit explores bias in hiring, helping individuals recognize how small biases in data or algorithms can lead to unfair outcomes. These games highlight the complexity of ethics in AI and the potential impact of biased decision-making.

3. Transparency in AI is essential to build trust and accountability in AI systems. It allows users to understand how AI models function, how decisions are made, and what data is being used. Transparency helps detect biases, ensure fairness, and protect privacy. It can be achieved by implementing explainable AI (XAI), which provides interpretable model outputs, clear documentation of the algorithms, and open communication about the limitations and risks of the AI system.

4. The purpose of intelligence ethics in AI development and use is to ensure that AI technologies are designed and implemented in a way that aligns with societal values and moral principles. Intelligence ethics seeks to reduce bias, and promote fairness, transparency, and accountability in AI systems. It guides responsible innovation, aiming to maximize the benefits of AI while addressing potential risks to human rights, equity, and privacy.

5. To mitigate the effects of bias in AI systems, several measures can be taken:
   - Use diverse and representative training datasets to prevent biased data from skewing results.
   - Regularly test and audit AI models for biases, ensuring ongoing fairness.
   - Employ explainable AI techniques to make AI decisions transparent and understandable.
   - Engage interdisciplinary teams to monitor AI ethics, bringing in diverse perspectives to assess potential biases and impacts.

C. Competency-based/Application-based questions:
   1. In the case of the drone's malfunction, the company responsible for its development and operation should be held accountable. Ethically, autonomous systems should prioritize human safety and be designed with fail-safes to minimize harm in unpredictable situations.

Transparency, accountability, and thorough testing should guide the development to ensure systems make ethical decisions and mitigate potential risks.

2. a. Both A and R are true and R is the correct explanation of A.

3. c. A is true but R is false.

Δi In Life (Page 424)

Do It Yourself

Δi Deep Thinking (Page 424)

Do It Yourself

Δi Lab (Page 425)

Do It Yourself

Class Activity (Page 425)

Do It Yourself

Δi Ready 8

Do It Yourself