Class 12

Answer Key

Part B-Subject Specific Skills

Artificial Intelligence (Ver. 3.0)

1. Python Programming II

Video Session (Page 151)

Do it yourself.

AI Reboot (Page 152)

1. `fillna(num)`
2. `isnull()`
3. `dropna()`
4. DataFrame
5. `read_csv()`

AI Task (Page 154)

1.
```
import pandas as pd
#Import the data
df = pd.read_csv("sales_data.csv")
```
2.
```
Perform basic inspection
print(df.head())  # Display the first five records
print(df.isnull())  # Display missing values in each column
# Fill missing numeric values
df['Price'].fillna(0, inplace=True)
df['Quantity'].fillna(0, inplace=True)
```
3.
```
Export cleaned data to a new CSV
df.to_csv("cleaned_sales_data.csv", index=False)
```

Exercise

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

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

B. 1. Data Science, and Analytics 2. Jupyter Notebook 3. rows and columns
4. index 5. analysing 6. shape
7. tail(n) 8. st.text_input()

SECTION B (Subjective Type Questions)

A. 1. NumPy uses Forward Indexing (Positive Indexing) and Backward Indexing (Negative Indexing).

2. One of NumPy's key features is its ndarray (N-dimensional array), which stores multiple values in a structured format. These arrays can have different ranks (dimensions)—a Rank 1 array is a simple one-dimensional list of numbers, while a Rank 2 array is a two-dimensional table of values.

3. Pandas offers Series (1D labeled array) and DataFrame (2D labeled data structure).

4. Streamlit is an open-source Python library used to build interactive web applications easily. Streamlit Community Cloud allows you to deploy and share Streamlit apps online directly from GitHub.

5. The head(n) function displays the first n rows of a DataFrame, defaulting to 5 if not specified.

B. 1. Pandas is a powerful Python library used for data analysis and manipulation. It provides tools to work with structured data, such as tables, spreadsheets, or databases. Let understand this with an example. Suppose we have a dataset containing information about various e-commerce orders, such as product category, price, discount applied, customer rating, number of units sold, and total revenue. Pandas allows us to load this dataset, explore trends, and analyse sales performance across different product categories. For example, we can use Pandas to calculate the average price of products in each category, find the best-selling items, and group data to see how discounts impact sales. If we want to visualize the revenue generated by different product categories, we can use Matplotlib, a popular plotting library in Python.

2. The steps to import and export data using Pandas are as follows:

Step 1: Use `read_csv("filename.csv")` to import CSV data into a DataFrame.

Step 2: Perform any data manipulation as needed.

Step 3: Use `to_csv("newfile.csv", index=False)` to export the DataFrame to a new CSV file.

3. **Linear Regression** is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. It helps predict the value of the dependent variable based on the independent variables.

   It is implement in Python by:

   - Importing necessary libraries such as `scikit-learn` and `pandas`.
   - Prepare the dataset, separating features (independent variables) and target (dependent variable).
   - Creating a Linear Regression model using `LinearRegression()` from `sklearn.linear_model`.
   - Training the model on the data using the `.fit()` method.
   - Making predictions using the `.predict()` method.
   - Evaluating the model's performance using metrics like R^2 score or Mean Squared Error (MSE).

4.

Method	Description	When to Use?
isnull()	Checks for missing values	To find out which values are missing
sum().sum()	Counts total NaN values	To understand the scale of missing data
dropna()	Removes rows with missing values	When missing data is small and won't affect analysis
fillna(value)	Fills missing values with default values	When missing data is important, and removing it isn't ideal

5. A Pandas DataFrame is a two-dimensional data structure that organizes data into labeled rows and columns, making it easy to manage and analyse. A Pandas DataFrame can be generated by importing datasets from various storage sources, such as SQL databases, CSV files, or Excel files. Additionally, a DataFrame can be constructed using lists, dictionaries, or a list of dictionaries, among other data structures.

6.

Numpy	Pandas
Focused on numerical and scientific calculations	Designed for data handling, analysis, and manipulation
Works with N-dimensional arrays (ndarrays)	Uses DataFrames and Series as main data structures
Requires all elements to be of the same data type	Supports mixed data types across different columns
Very fast for large-scale numerical tasks	Efficient for labelled, tabular data and works well with large datasets

C. Competency-based/Application-based questions.

1. Use fillna() to fill missing marks in the 'Math' column (e.g., with mean or median).
2. Use dropna() to remove rows with missing feedback.

Assertion and Reasoning Questions.

3. d.
4. a.

Statement-based Questions

5. d.
6. a.

 **Ai Lab** (Page 170)

Do it yourself.

 **Ai In Life** (Page 170)

Do it yourself.

 **Ai Deep Thinking** (Page 170)

Do it yourself.

Ai Ready (Page 171)

Do it yourself.

2. Data Science Methodology: An Analytic Approach to Capstone project

Video Session (Page 178)

Do it yourself.

 **Ai Reboot** (Page 193)

1. Step 1: Business Understanding – The foundation of every AI Project

   Step 2: Analytic Approach

   Step 3: Understanding Data requirements in a Project

   Step 4: Data Collection

Step 5: Understanding the Data we collected

Step 6: Data Preparation

Step 7: AI Modeling

Step 8: Evaluation of an AI Model

Step 9: Deployment of an AI Model

Step 10: Feedback

2. Data preparation

Video Session (Page 196)

Do it yourself.

Video Session (Page 198)

Do it yourself.

Ai Task (Page 204)

Actual (A)	Predicted (P)	(A - P)	$(A - P)^2$
105	115	-10	100
130	120	10	100
140	135	5	25
160	155	5	25
180	175	5	25
200	195	5	25
220	230	-10	100
250	240	10	100
270	275	-5	25
300	290	10	100

Step 1: Add up all $(A - P)^2$ values

Sum of $(A - P)^2$ = 100 + 100 + 25 + 25 + 25 + 25 + 100 + 100 + 25 + 100 = 625

Step 2: Calculate MSE (Mean Squared Error)

MSE=625/10=62.5

Step 3: Calculate RMSE (Root Mean Squared Error)

RMSE=√62.5 = 7.91

Mean Squared Error (MSE) = 62.5

Root Mean Squared Error (RMSE) ≈ 7.91

Exercise

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

A.

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

B.

1. Descriptive	2. risks	3. Simulation	4. real-time
5. Kaggle, data.gov	6. Modelling	7. Model	8. Deployment
9. Recall	10. 0		

C.

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

SECTION B (Subjective Type Questions)

A.

1. The key objective is to identify the business problem clearly and formulate a structured method to solve it. It involves analysing customer needs using frameworks like 5W1H and Design Thinking to define objectives aligned with business goals.
2. Selecting the correct analytical approach ensures that data is analysed effectively to find meaningful insights. It guides the choice of the right method like regression, classification, clustering, or recommendation, depending on the project needs.
3. The four types are:
   - Descriptive Analytics
   - Diagnostic Analytics
   - Predictive Analytics
   - Prescriptive Analytics

   Each type plays a role from summarising past data to recommending actions for future outcomes.
4. Two techniques used in Predictive Analytics are:
   - Regression (finding relationships between factors)
   - Classification (categorising data into groups)

   Both techniques help forecast future trends based on historical data.
5. The key steps are cleaning data by removing missing values and errors, combining data from multiple sources, transforming raw data into useful variables, and creating new features through Feature Engineering to improve model accuracy.

6. A Confusion Matrix evaluates a classification model by comparing actual outcomes with predicted outcomes. It helps identify correct predictions (True Positives, True Negatives) and errors (False Positives, False Negatives) to measure model performance.
7. Accuracy measures the overall correctness of a model by dividing the number of correct predictions (True Positives + True Negatives) by the total number of predictions. Higher accuracy means the model is performing well.
8. The F1-score balances Precision and Recall, ensuring a model is accurate and does not miss important predictions. It is important when both False Positives and False Negatives need to be considered, providing a better evaluation.
9. RMSE measures how much predicted values differ from actual values in regression models. It shows the spread of errors and is in the same unit as the target variable. Lower RMSE indicates better model performance.
10. Yes, Feedback is necessary as it helps improve the model based on real-world results. It ensures that the model stays accurate and reliable by refining it through user feedback, performance reports, and continuous updates.

B. 1. Data can be classified into three main types:

- **Structured Data** – Organised in a fixed format, such as tables (e.g., customer databases).
- **Semi-Structured Data** – Has some organisation but is not fully structured, like emails or XML files.
- **Unstructured Data** – Data that does not follow a set format. For example, social media posts, images, or videos.

2. Secondary data refers to pre-existing information that has already been collected, processed, and stored by others. It is valuable for research and analysis without requiring direct data collection. Common sources include:

- **Books, Journals, & Research Papers** – Academic studies, historical records, and scholarly articles.
- **Websites & Online Databases** – Platforms like Google Scholar, World Bank Open Data, UNICEF, Kaggle, and data.gov provide readily available datasets.
- **Internal Company Databases** – Past sales records, employee performance reports, and financial statements used for business insights.
- **Social media & Web scraping** – Extracting data from platforms like Twitter, Instagram, and online reviews to analyse trends.

For example, a business analyses sales data from the past five years to forecast future market trends.

3. **DBAs (Database Administrators)** manage and organise data by designing database structures, setting data standards, and ensuring security and accuracy. They keep the database safe, consistent, and working well.

Programmers create tools like apps and forms to collect data. They write code to validate, clean, and integrate data from different sources.

Accuracy and reliability are important because AI systems depend on clean, correct data to make good decisions. Incorrect data can lead to wrong results, loss of trust, and poor model performance. Accuracy ensures correct decision-making and reliability ensures consistent results over time, both critical for trustworthy AI solutions.

4. The Data Understanding stage involves evaluating the collected dataset to ensure it is relevant, complete, and suitable for solving the given problem. This process helps determine whether the data truly represents the issue at hand or if additional collection or adjustments are needed.

   This step is important because:

   - Ensures that the collected data is relevant to the problem being addressed.
   - Helps identify any missing, inaccurate, or irrelevant data before proceeding to the next stage.
   - Allows for assessing whether additional data collection or corrections are needed.

5. Data preparation is a crucial step in any data science project as it ensures that the data is clean, structured, and ready for analysis. Without proper data preparation, Machine Learning models may produce inaccurate results.

   - **Most time-consuming step** – Data preparation takes up 80% of the total time in a data science project.
   - **Common for different problems** – Many industries follow similar data preparation techniques.
   - **Automation speeds up the process** – Using AI-powered tools, data preparation can be faster and more efficient by reducing manual work.

   Automating certain data preparation steps in advance can speed up the process by minimising ad hoc preparation time. Today's high-performance, massively parallel systems and analytics capabilities where data is stored allow data scientists to prepare data more easily and quickly using very large datasets.

6.

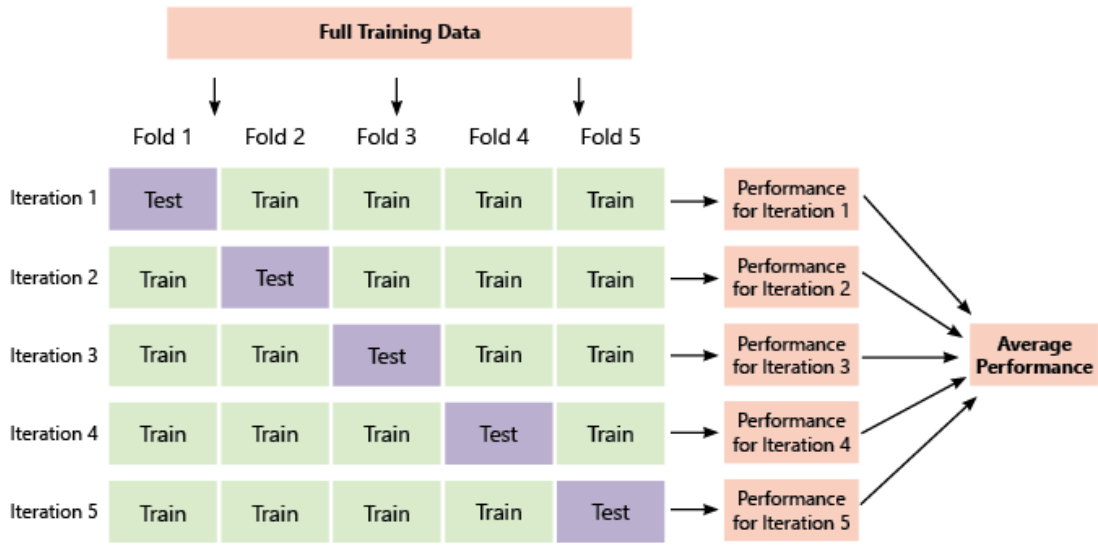
Descriptive modelling	Predictive modelling
Summarizes past data using techniques like descriptive analytics.	Uses historical data to predict future outcomes using regression or classification techniques.
It helps understand historical trends.	Helps in forecasting and decision-making.
For example, a supermarket manager can analyse sales data from the past six months to determine the most popular products.	For example, a mobile company can analyse past customer complaints and usage patterns to predict which customers are likely to stop using their service.

7. In k-fold cross-validation, data is split into k parts. The model is trained on k-1 parts and tested on the remaining part. This is repeated k times with different folds, and the average performance is measured. It ensures robust model evaluation.

   For example, if we choose k = 5, the dataset is divided into five equal folds, each containing 20% of the data.

   - In the first experiment, we train the model on four folds and test it on the remaining fold.
   - In the second experiment, we hold out a different fold for testing and train on the other four.
   - This continues until each fold has been used as a test set once.

   Since every data point is used for testing at some point, k-fold cross-validation provides a fair and accurate estimate of the model's performance. By the end of the process, you have 5 different performance measures, which you can average to get a more reliable estimate of the model's accuracy.



8. **Similarities:**

   Both methods are used for model evaluation.

   Differences:

Train-Test Split	Cross Validation
Train-test split is a quick way to check a model's accuracy.	Cross-validation gives a more thorough evaluation by testing the model multiple times on different data segments.
It divides the data into two parts – one for training and one for testing.	It divides the dataset into multiple smaller parts (folds).

The training set is used to teach the model, while the test set checks how well it performs on new data.	The model is trained on some folds and tested on others, repeating the process several times. This ensures that every data point is used for both training and testing at different stages.
Mainly used for large datasets.	It is beneficial for smaller datasets.

9. The step related to constructing the data set is Data Collection.

   Data collection is a systematic process of gathering information through observations, measurements, or surveys. During this phase, data requirements are reviewed, and decisions are made on whether additional data is needed based on project objectives. With advancements in high-performance database systems, data scientists can efficiently manage and analyse large datasets.

   It gathers information from primary sources like surveys and sensors, and secondary sources like databases. It ensures sufficient and relevant data for analysis and modeling.

10. Data Preparation involves cleaning errors, handling missing values, integrating data from multiple sources, and engineering new features.

    The **Data Preparation** stage includes all the activities needed to organise and process raw data before using it in the modelling step. This process transforms data into a format that makes it easier to work with and analyse. The steps in data preparation are as follows:

    - **Cleaning Data** – Removing missing values, duplicate entries, and formatting data correctly.
    - **Combining Data** – Merging information from different sources like tables, databases, and files.
    - **Transforming Data** – Converting raw data into meaningful input variables.

    These tasks are time-consuming but essential to ensure that the data is of high quality, directly impacting model success.

C. Competency-based/Application-based questions.

1. Confusion Matrix Calculations:
   - Precision: 50 / (50 + 5) = 90.9%.
   - Recall: 50 / (50 + 10) = 83.3%.
   - F1 Score: 2*(0.909*0.833)/(0.909+0.833) =0.8688 ≈ 87%.
   - Accuracy: (50 + 35) / 100 = 85%
2. Use **cross-validation**.

   Reason: It gives more reliable results by using the small dataset more efficiently and reducing overfitting.
3. Use a **predictive analytics model**.

   Reason: It helps forecast maintenance using sensor data like battery, temperature, and mileage.

Assertion and Reasoning Questions.

4. c.
5. a.

Statement–based questions

6. c.
7. a.

 **Ai Lab** (Page 217)

Do it yourself.

 **Ai Deep Thinking** (Page 217)

Do it yourself.

Ai Ready (Page 218)

Do it yourself.

3. Making Machines See

 **Ai Task** (Page 220)

Do it yourself.

 **Ai Reboot** (Page 222)

1. Computer Vision is faster and more accurate because machines can process thousands of images in seconds, never get tired, and can identify patterns or tiny differences that human eyes may miss.
2. Higher bit depth is important because it allows an image to represent more colors and finer details, improving the quality and making the images more realistic and accurate during processing.

 **Ai Task** (Page 227)

Do it yourself.

 **Ai Reboot** (Page 238)

1. Depth
2. Preprocessing

3. Feature extraction
4. Multiple

Video Session (Page 240)

Do it yourself.

Ai Task (Page 242)

Do it yourself.

Video Session (Page 252)

Do it yourself.

Exercise

Unsolved Questions

SECTION A (Objective Type Questions)

Ai Quiz

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

B. 1. Features from Accelerated Segment Test
2. Gaussian Blur
3. edge detection
4. motion detection
5. deep cognitive reasoning
6. high-level

SECTION B (Subjective Type Questions)

A. 1. **Resizing images** means changing the dimensions of an image, usually to fit a particular requirement or reduce computational load. It can be done by increasing or decreasing the number of pixels without significantly altering the content.

2. **Histogram Equalization** is a preprocessing technique used to improve image contrast by redistributing the pixel intensity values. This helps in highlighting key features in low-contrast images.

3. **Detection and segmentation** identify and separate different objects or regions in an image. Detection draws bounding boxes around objects, while segmentation assigns labels to each pixel, helping in more precise object recognition.

4. **OpenCV** (Open Source Computer Vision Library) is a popular library in Python and C++ that provides tools for image and video processing, helping machines to detect, recognize, and understand visual data.

5. An image can be **converted to Grayscale** using functions like cv2.cvtColor(image, cv2. COLOR_BGR2GRAY) in OpenCV, which simplifies the image to shades of gray, useful for reducing complexity in processing.

6. **Classification** identifies what an object is (e.g., cat, dog), while **detection** identifies what and where it is in the image using bounding boxes.
7. Two algorithms used for object detection are:
   - YOLO (You Only Look Once)
   - SSD (Single Shot Detector)
8. In **single object detection**, the system identifies one object by processing an image, extracting features, classifying it, and drawing a bounding box around it using algorithms like YOLO or SSD.
9. Applications of Computer Vision include:
   - Facial recognition
   - Medical imaging (e.g., tumor detection)
   - Surveillance
   - Automated checkout in retail
10. **Noise reduction** improves image quality by eliminating unwanted distortions. This ensures that only the relevant features remain, enhancing the accuracy of object detection and recognition.

B. 1. Some challenges in computer vision are as follows:
   - **Complex Reasoning:** CV struggles with understanding context, emotions, and predicting actions due to limited cognitive reasoning.
   - **Image Acquisition Issues:** Variations in lighting, motion blur, occlusions, and noise affect image quality and analysis.
   - **Privacy and Security:** Risks include unauthorized data use, biased models, and adversarial attacks that deceive systems.
   - **Fake and Manipulated Content:** Deepfakes and synthetic images cause misinformation and trust issues.
   - **Lack of Explainability:** Deep learning models often act like "black boxes," making it hard to explain decisions.
   - **Multimodal Integration:** Combining images with other data types (audio, sensors) is complex and challenging.

2. Five applications of CV:
   - Healthcare and medical imaging
   - Facial recognition
   - Machine inspection and quality control
   - 3D model construction and reconstruction
   - OCR for document digitisation

3. High-level processing is the final stage of Computer Vision that focuses on understanding and interpreting visual content to facilitate decision-making. At this stage, Computer Vision systems go beyond basic object detection— they recognise, analyse, and infer contextual meaning from images, enabling intelligent decision-making based on visual data.

   Key aspects of high-level processing are as follows:

   1. Object recognition: High-level processing enables computers to identify and classify objects within an image.

      Example: In a security surveillance system, the system can distinguish between a car, a person, or an animal, enabling automated threat detection or monitoring.

   2. Scene understanding: This involves analysing the entire image to interpret spatial relationships and contextual meaning.

      Example: In autonomous vehicles, high-level processing evaluates road environments, detecting elements such as lanes, traffic signals, pedestrians, and nearby vehicles to assist in navigation and decision-making.

   3. Context analysis: Beyond recognising objects, high-level processing identifies relationships between objects and their arrangement within a scene.

      Example: In retail automation, a system can analyse a grocery store image and determine that fruits placed in baskets are ready for checkout, while shelves contain canned goods, improving inventory management.

4.

Aspect	Single Object Tasks	Multiple Object Tasks
Definition	Tasks focusing on detecting or analysing one object in an image or video.	Tasks involving detection, recognition, or tracking of several objects simultaneously.
Complexity	Relatively simpler due to the focus on a single object.	More complex because of multiple objects and interactions among them.
Examples	Face recognition, single object tracking, barcode scanning.	Crowd counting, traffic monitoring, multi-object tracking in sports.
Challenges	Handling variations of one object (pose, scale, occlusion).	Differentiating and tracking multiple overlapping or interacting objects.
Applications	Access control, single product inspection, license plate recognition.	Surveillance, autonomous driving, retail analytics.

5.

Resizing	Cropping
Changes the entire image size to fixed dimensions.	Cuts out a specific part of the image.
Makes all images uniform for machine learning models.	Focuses on important areas by removing background.
May cause some distortion or loss of detail.	Keeps details but reduces image area.

6. Real-world applications of CV include:
   - Healthcare for early disease detection.
   - Agriculture for crop health monitoring.
   - Industrial manufacturing for quality control.
   - Security systems for threat detection.
   - Retail for customer experience enhancement.

C. Competency-based/Application-based questions.

1. Higher resolution images contain more pixels per unit area, resulting in finer detail and better clarity. When digitizing, this ensures that the visual information is retained accurately. High-resolution images maintain quality when zoomed in or printed, whereas low-resolution images may appear pixelated or blurry on screens or paper.
2. i. Motion detection or Edge detection to identify unusual movement or boundaries in blind spots.
   ii. Activity recognition using texture analysis and motion features to detect crowd behaviour.
   iii. **Colour-based feature extraction** (to detect the red scarf) combined with **facial recognition** using CNNs.

Assertion and Reasoning Questions.

3. a.
4. a.

Statement–based questions

5. a.
6. a.

Ai In Life (Page 261)

Do it yourself.

Ai Deep Thinking (Page 261)

Do it yourself.

Ai Lab (Page 261)

Do it yourself.

Ai Ready (Page 263)

Do it yourself.

4. AI with Orange Data Mining Tool

Ai Reboot (Page 267)

1. Data mining is important because it extracts valuable insights, patterns, and trends from vast amounts of data. It helps organisations uncover hidden relationships and meaningful information, enabling them to make data-driven decisions, predict future trends, and optimise business strategies using statistical techniques and machine learning algorithms.
2. Orange Data Mining is beneficial for researchers as it allows data exploration, hypothesis testing, and pattern recognition without extensive programming skills. It supports handling large datasets, text mining, and network analysis, aiding researchers in fields like healthcare, social sciences, and bioinformatics.

Ai Reboot (Page 271)

1. CSV File Import: Loads CSV files with flexible settings for delimiters and encoding.
2. Preprocess: Allows users to select and combine various preprocessing techniques in a single interface.
3. Mosaic Display: Shows associations between categorical variables.
4. Gradient Boosting: Boosted trees for better predictive performance.
5. ROC Analysis: Shows ROC curves and AUC scores.
6. DBSCAN: A density-based clustering method.

Ai Task (Page 282)

Do it yourself.

Video Session (Page 282)

Do it yourself.

Ai Task (Page 287)

Do it yourself.

Ai Task (Page 292)

Do it yourself.

Exercise

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

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

B. 1. open-source, machine-learning 2. modular 3. data loading
4. Data Info 5. Create Instance 6. FreeViz 7. relationships
8. Data Science

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

SECTION B (Subjective Type Questions)

A. 1. **Orange tool's widgets support tasks** such as data loading, transformation, visualisation, modelling, evaluation, and clustering. For example, File loads data, Scatter Plot visualises relationships, Tree builds models, and Confusion Matrix evaluates results.

2. **Default widgets** in Orange are built-in tools that provide essential functions such as data preprocessing, transformation, visualisation, modelling, evaluation, and unsupervised learning. They are available under various categories and are the core components used in workflows.

3. Evaluation widgets help in assessing model performance by providing statistical metrics and validation techniques. Users can evaluate models using cross-validation, view confusion matrices, generate ROC curves, and assess performance through various scoring methods. Tools like permutation plots and parameter tuning allow for fine-tuning models, while calibration plots provide insights into probability estimates.

4. **Add-ons in Orange** are extensions that provide specialised widgets and features beyond the default set. They enable advanced functionalities like text mining, image analytics, and network analysis, and can be added via Options → Add-ons.

5. The **Corpus Viewer** widget helps users view and navigate textual data. It highlights searched words in context, enabling quick exploration and understanding of the text within a selected corpus, such as Grimm Tales in text mining workflows.

B. 1. The **Corpus Viewer widget** allows users to examine text from a corpus in a readable, scrollable window. It highlights selected words using a regular expression filter. It is particularly useful for identifying key themes, browsing occurrences of specific words, and gaining preliminary insights into textual data before performing further text mining operations such as word cloud generation or preprocessing.

2. The **Preprocess Text widget** is used to clean and prepare text data for analysis. It carries out multiple tasks including converting text to lowercase, tokenizing, removing stopwords, normalising text, and lemmatization. These processes are crucial for improving model accuracy by reducing noise and standardising the input text, making it more meaningful for tasks like classification and topic modelling. Text preprocessing plays a crucial role in any text mining workflow, as it shapes the data that machine learning models and analytical algorithms will process.
3. **Visualize Widgets** provide graphical representations of data to help users understand distributions, trends, and relationships. Examples include Scatter Plot (shows feature relationships), Box Plot (identifies outliers), Mosaic Display (shows associations in categorical data), and Heat Map (visualises correlations). These visual tools aid in better data exploration and model interpretation.
4. **Beneficiaries of Orange Data Mining** include:
   - Data Analysts for no-code modelling.
   - Researchers for text mining and exploration.
   - Educators and Students for interactive learning.
   - Business Professionals for predictive analytics.
   - Developers via open-source contributions.

   Each group uses Orange to simplify data analysis, modelling, and AI tasks.
5. **Steps to install Image Analytics widget in Orange Tool:**
   - Open a blank canvas
   - Go to **Options** → **Add-ons**
   - In the dialog box, select **Image Analytics**
   - Click **OK**
   - Restart Orange
   - The Image Analytics widget will appear in the widget list

C. Competency-based/Application-based questions.

1. Nitin should install the Text add-on in Orange Data Mining. This provides widgets like Corpus, Preprocess Text, Word Cloud, and Sentiment Analysis, enabling him to analyse text data effectively and classify reviews based on sentiment.
2. Vikram should use the Box Plot widget. This widget displays the distribution of numerical data and identifies outliers. It combines statistical summary with visual representation, making it ideal for exploring variations in values.

Assertion and Reasoning Questions.

3. c.
4. a.

Statement–based questions

5. a.
6. c.

 **Ai In Life** (Page 299)

Do it yourself.

 **Ai Deep Thinking** (Page 300)

Do it yourself.

 **Ai Lab** (Page 300)

Do it yourself.

Ai Ready (Page 302)

Do it yourself.

5. Introduction to Big Data and Data Analytics

 **Ai Reboot** (Page 308)

1. Big Data refers to extremely large and complex datasets that cannot be managed or analysed using traditional data-processing tools. It includes structured, semi-structured, and unstructured data and requires advanced techniques to extract meaningful insights.
2. Two characteristics of data are as follows:
   - Volume: Refers to the vast amount of data generated.
   - Velocity: Refers to the speed at which data is created and processed.
3. Structured Data

   Unstructured Data

   Unstructured Data

   Semi-structured Data
4. Big data mainly comes from three sources:
   - **Transactional Data:** Sales records, banking transactions, and purchase history.
   - **Machine Data:** Logs from sensors, GPS tracking, and automated systems.
   - **Social Data:** Posts, likes, comments, and interactions on social media.
5. Unstructured Data is widely used in AI, such as images, videos, social media content, and natural language.

Ai Reboot (Page 310)

1. b. By enabling predictive analytics for disease detection
2. b. By using AI algorithms to detect unusual spending patterns
3. a. To suggest personalised product recommendations
4. a. By predicting when machines will fail before breakdowns occur
5. a. By tracking network performance and predicting customer churn

Ai Task (Page 311)

Do it yourself.

Video Session (Page 311)

Do it yourself.

Video Session (Page 316)

Do it yourself.

Ai Task (Page 317)

Do it yourself.

Ai Task (Page 329)

Do it yourself.

Video Session (Page 331)

Do it yourself.

Ai Reboot (Page 331)

Steps in the Working Process of Big Data Analytics

Step 1: Data Collection

Step 2: Data Storage

Step 3: Data Cleaning and Processing

Step 4: Data Analysis and Interpretation

Exercise

Unsolved Questions

SECTION A (Objective Type Questions)

Ai Quiz

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

B. 1. Velocity 2. Veracity 3. innovation 4. Variability 5. Big Data Analytics
6. Cloud computing 7. data stream 8. heat map

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

SECTION B (Subjective Type Questions)

A. 1. Governments use Big Data to analyse public feedback, monitor service delivery, predict demands, and improve efficiency in areas like healthcare, transportation, and urban planning. It helps in evidence-based policy making and proactive governance.

2. Businesses may face challenges such as lack of skilled personnel, high infrastructure costs, data privacy concerns, and the complexity of integrating Big Data with existing systems.

3. 'Value' signifies the usefulness of data in generating meaningful insights. Without value, even large volumes of data are not beneficial. Data must be relevant, accurate, and timely to support decision-making.

4. Data Analytics is the process of examining large and varied data sets to uncover hidden patterns, correlations, trends, and insights that can help organisations make informed decisions.

5. The main goal of Big Data Analytics is to extract actionable insights from massive datasets to improve operations, enhance customer experience, optimise resources, and support strategic decisions.

6. Challenges include managing data volume and quality, ensuring data privacy, securing skilled talent, integrating multiple data sources, and implementing scalable storage and processing systems.

7. The rise of Big Data Analytics is driven by four major technological advancements – Moore's Law, Mobile Computing, Social Networking and Cloud Computing.

8. Moore's Law predicts that the number of transistors on a microchip doubles approximately every two years, leading to exponential growth in computing power and data generation.

9. Batch Processing handles large volumes of data in groups at scheduled times, while Stream Processing handles real-time data as it is generated, suitable for applications needing immediate insights.

10. K-means is an unsupervised machine learning algorithm that groups data into 'k' clusters based on similarity. It minimises variance within clusters and is used in market segmentation and pattern recognition.

B. 1. Big Data Analytics involves four main steps:

1. **Gather Data** from various sources like cloud storage, apps, and IoT devices.
2. **Process Data** using batch processing for large datasets or stream processing for real-time data.
3. **Clean Data** by fixing errors, removing duplicates, and correcting formats to ensure quality.

4. **Analyse Data** with advanced tools to find patterns and insights that help businesses make better decisions.

2. **Unstructured data** is data without a fixed format, like text, images, videos, or social media posts. Unlike **structured data**, which is organised in tables (like spreadsheets), unstructured data is harder to store and analyse.

   Technologies used to analyse unstructured data include Natural Language Processing (NLP), Computer Vision, Deep Learning, and Big Data tools like Hadoop.

   Examples: Emails, photos, videos, audio files, and social media content.

3. Big Data Analytics offers significant advantages:

   Real-Time Intelligence: Big Data Analytics enables organisations to process vast amounts of data as it is generated, providing real-time insights. This allows businesses to make swift decisions, respond instantly to market shifts, and seize emerging opportunities.

   Better Decision-Making: By uncovering hidden patterns and trends, Big Data Analytics empowers decision-makers with actionable insights. This enhances strategic planning across supply chain management, e-commerce, operations, and business growth.

   Cost Efficiency: Identifying inefficiencies and waste, Big Data Analytics helps streamline operations and reduce expenses. Predictive analytics also enables better resource allocation, preventing costly missteps.

   Enhanced Customer Engagement: Understanding customer behaviour, preferences, and sentiment allows businesses to personalise marketing strategies. This leads to improved customer satisfaction and stronger brand loyalty.

4. Three Types of Big Data Analysis:

   - **Descriptive Analytics**

     It explains **what happened** by summarising past and present data using statistics like mean, median, and frequency.

     Example: A retail company analysing monthly sales reports.

     Tools: Excel, Tableau, SQL.

   - **Diagnostic Analytics**

     It explains **why something happened** by finding causes through root cause analysis and correlation.

     Example: A hospital studying why patient readmission rates increased.

     Tools: Python, R, Google Analytics.

   - **Predictive Analytics**

     It forecasts **what will happen** using historical data and machine learning to predict future trends.

     Example: An e-commerce platform predicting which customers will buy a product.

     Tools: Regression models, classification algorithms, AI models.

5.

Small Data	Big Data
Small, easy to understand and analyse.	Very large and complex datasets.
Used for quick, simple decision-making.	Requires advanced tools and techniques to analyse.
Examples: Daily sales, attendance records.	Examples: Online purchases, sensor data, social media.
Does not need specialised technology.	Needs specialised storage and processing systems.

6. Data Stream Mining involves:
   - **Continuous Flow**

     Data arrives in real time and must be processed instantly.
   - **High Volume and Velocity**

     Data streams generate vast amounts of data at high speeds.
   - **One-Pass Processing**

     Data is processed once and discarded as the storage is limited.
   - **Concept Drift**

     Data patterns may change over time, requiring adaptive models.

   Additionally, mining data streams allows organisations to **process, analyse, and act on high-speed data in real-time**, which is essential for fraud prevention, predictive analytics, and smart decision-making.
7. The future of Big Data Analytics is very promising, with rapid growth expected globally, including in India. It will help businesses make smarter decisions, improve efficiency, and drive innovation. Key technological advancements shaping its growth include:
   - **Real-Time Analytics:** Enables instant data processing for immediate decision-making.
   - **Advanced Predictive Models:** Uses AI and Machine Learning to forecast trends accurately.
   - **Quantum Computing:** Will significantly speed up data processing and solve complex problems.

   These developments will transform industries, create new job opportunities, and continue to drive digital innovation.
8. As businesses increasingly rely on data for decision-making, efficiency, and customer engagement, demand for professionals skilled in Big Data Analytics is growing. Here are key career paths in the field:
   - **Data Scientist:** Analyses complex data using Machine Learning and predictive modelling to uncover insights that drive business decisions.
   - **Data Analyst:** Extracts meaningful trends from data using statistical techniques to support strategic planning.
   - **Data Engineer:** Develops and maintains data infrastructure, ensuring efficient processing and management of large datasets.

- **Machine Learning Engineer:** Designs and implements algorithms that enable systems to learn from and make predictions on data.
- **Business Intelligence (BI) Analyst:** Transforms raw data into actionable insights through reports and visualisations for business stakeholders.
- **Data Visualisation Specialist:** Creates graphical representations of data to enhance comprehension and decision-making.

C. Competency-based/Application-based questions.

1. Unstructured Data — as it includes diverse formats like social media comments, blog posts, and news articles.
2. a. Call logs, signal strength, user locations, device types, customer complaint logs, feedback forms and network traffic data.

   b. By analysing geo-tagged data and user complaints in real time to pinpoint high-failure zones. Identify network congestion or hardware faults by analysing **traffic spikes or signal quality degradation**.

   c. Decision Trees, Neural Networks, and Time Series models can forecast network congestion.

   d. By enabling proactive resolution of network issues, improving service delivery, and reducing call drops. Enhanced **user experience** leads to better customer retention and positive sentiment.

Assertion and Reasoning Questions.

3. b.
4. c.

Statement–based questions

5. c.
6. a.

Ai In Life (Page 341)

Do it yourself.

Ai Deep Thinking (Page 342)

Do it yourself.

Ai Lab (Page 342)

Do it yourself.

Ai Ready (Page 343)

Do it yourself.

6. Understanding Neural Networks

Ai GAME 01 (Page 345)

Do it yourself.

Video Session (Page 345)

Do it yourself.

Ai Reboot (Page 348)

A. 1. A Neural Network is a system made up of interconnected neurons. It replicates the function of human neurons by forming an Artificial Neural Network (ANN) that adjusts to inputs and produces results without external modifications, similar to how the human brain works.

2. The three layers of a Neural Network are the **Input Layer**, **Hidden Layer**, and **Output Layer**.

B. 1. weight

2. hidden

3. themselves without requiring explicit programming

Video Session (Page 348)

Do it yourself.

Ai GAME 02 (Page 352)

Do it yourself.

Video Session (Page 354)

Do it yourself.

Video Session (Page 355)

Do it yourself.

Ai Task (Page 357)

Do it yourself.

Ai Task (Page 359)

Do it yourself.

Ai Task (Page 362)

Do it yourself.

Ai GAME 03 (Page 363)

Do it yourself.

Video Session (Page 364)

Do it yourself.

Ai GAME 04 (Page 365)

Do it yourself.

Ai Task (Page 367)

Do it yourself.

Exercise

Unsolved Questions

SECTION A (Objective Type Questions)

Ai Quiz

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

B. 1. Output Layer 2. Tanh function 3. hidden
4. Deep Neural Network 5. Bias 6. epochs
7. non-linearity 8. threshold logic unit 9. Pooling
10. Perceptron

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

SECTION B (Subjective Type Questions)

1. An Artificial Neural Network (ANN) is a system of interconnected neurons that mimic the function of the human brain. It processes inputs, learns from them, and produces outputs by adjusting weights and biases without requiring explicit programming.
2. The Output Layer is the final layer of a neural network. It receives the processed information from the hidden layers and makes decisions or predictions based on the learned patterns.
3. Deep Learning is a subset of Machine Learning that uses Neural Networks with many hidden layers. These networks learn from large volumes of data and improve performance by adjusting weights and biases across multiple layers.
4. Neural Networks are widely used in applications such as smart compose, spam filtering in emails, voice recognition, facial recognition, medical diagnosis, and stock market prediction.
5. In a neural network, inputs are multiplied by weights and summed. This weighted sum is passed through an activation function, helping the network learn patterns by adjusting the importance of inputs during training.

6. The three main layers of a Convolutional Neural Network (CNN) are the Convolutional Layer, Pooling Layer, and Fully Connected Layer. These help in feature extraction, data reduction, and final classification.

B. 1. Neurons, also known as nodes, are the fundamental building blocks of a Neural Network. They receive inputs from other neurons or external sources, such as numbers or images. Each neuron calculates a weighted sum of its inputs, applies a rule known as an activation function, and then generates an output, which is transmitted to the next layer of neurons. This process allows the network to learn patterns and make predictions based on the given data.

For example, imagine a neuron analysing whether an email is spam. It receives inputs such as keywords, the sender's address, or the subject line. Each of these inputs is assigned a weight based on its importance, and the neuron calculates a weighted sum. It then applies an activation function to determine whether the email is likely spam, passing the result to the next layer of neurons for further analysis. This process helps the Neural Network classify emails accurately.

2. Activation functions are like on/off switches for neurons. They decide if a neuron should send a signal or stay quiet. Activation functions in Neural Networks act as decision-makers for each neuron, determining whether it should activate (send a signal) based on the input it receives. After computing the input, the activation function applies a rule to determine the output, adding flexibility and enabling the network to learn complex patterns. These functions introduce non-linearity to the model, allowing it to capture intricate relationships in data. Common activation functions include the Sigmoid Function (which provides a smooth transition), the Tanh Function (which balances the output), and ReLU (Rectified Linear Unit), which turns off for negative inputs.

For example, a self-driving car uses activation functions to decide whether to stop, turn, or accelerate based on input from sensors, such as detecting obstacles, interpreting traffic signals, or recognising lane markings.

3. A Neural Network processes data through the following steps:

1. **Input Data Processing**

The network receives input (e.g., image pixels or text), where each input neuron represents a feature.

2. **Weighted Sum Calculation**

It computes the weighted sum of inputs using the formula:

$\sum (Wi \times Xi) + b$

3. **Activation and Layer Propagation**

The weighted sum is passed through an activation function (like ReLU or Sigmoid), and the data flows through hidden layers to extract features.

4. **Output Generation**

   The final layer produces the result—either a label (e.g., "cat") or a value (e.g., price)—based on the processed data.

5. **Error Calculation**

   A loss function compares the predicted output with the actual result to measure error.

6. **Backpropagation and Weight Update**

   The network adjusts weights and bias to reduce error using backpropagation.

7. **Training Iterations (Epochs)**

   These steps repeat over multiple epochs to improve accuracy.

Once trained, the network can predict new, unseen data in areas like AI, computer vision, and NLP.

4. A Neural Network works by passing data through a series of interconnected nodes or neurons arranged in layers. Each node acts as a small processing unit that receives inputs, multiplies each input by a specific weight, adds a bias (a constant value that shifts the result), and then applies an activation function to produce an output. This output is then passed forward to the next layer of nodes in what is known as a feedforward network. The bias is always present and plays a key role in adjusting the position of the activation function, allowing the network to learn more effectively. During training, the values of weights and bias are continuously updated to improve the model's predictions. A threshold is often used as a decision boundary to determine whether the node should be activated or not. Through this layered process of computation and learning, the neural network is able to recognise patterns, make decisions, and generate predictions based on the input data.
5. Recurrent Neural Networks (RNNs) are designed to handle sequential data, where the order of information matters. They feature feedback loops that allow information to persist across time steps, enabling the network to remember previous inputs. This makes RNNs particularly effective for tasks that unfold over time. These feedback connections create a loop where data continuously flows. If the prediction is incorrect, backpropagation adjusts the network by making small changes using the learning rate, gradually improving accuracy over iterations.

   Applications: RNNs are used for music and speech generation; robotics & autonomous systems; time series forecasting like stock prices or weather patterns.

   For example, Google Translate? RNNs power its ability to read and translate entire sentences while keeping the meaning intact. Similarly, virtual assistants like Alexa use RNNs to understand and respond to spoken commands.
6. Neural Networks have transformed industries by improving efficiency, enabling automation, and supporting decision-making in healthcare, finance, and education. They offer personalised services and contribute to economic growth.

However, they also pose risks like data privacy concerns, algorithmic bias, and job displacement. To address these, it is important to enforce strong data protection laws, use diverse datasets, conduct bias audits, and invest in reskilling the workforce.

Ensuring ethical development, transparency, and regulations will help maximise benefits while minimising harm to society.

C. Competency-based/Application-based questions.

1. College Decision (Inputs and Weights Example):
   - Inputs: Academic reputation, campus facilities, tuition fees
   - Weights (Example): Academic reputation = 0.5, Campus facilities = 0.3, Tuition fees = 0.2
   - Formula:

     Weighted Sum = (Academic reputation × 0.5) + (Campus facilities × 0.3) + (Tuition fees × 0.2)

     Use numeric scores for each factor and calculate accordingly.
2. Restaurant Choice (Inputs and Weights Example):
   - Inputs: Food quality, ambience, distance
   - Weights (Example): Food quality = 0.4, Ambience = 0.4, Distance = 0.2
   - Formula:

     Weighted Sum = (Food quality × 0.4) + (Ambience × 0.4) + (Distance × 0.2)

     Assign values to each factor and compute the weighted sum for comparison.

Assertion and Reasoning Questions.

3. a.
4. a.

Statement–based questions

5. c.
6. a.

Ai In Life (Page 381)

Do it yourself.

Ai Deep Thinking (Page 381)

Do it yourself.

Ai Lab (Page 382)

Do it yourself.

Ai Ready (Page 384)

Do it yourself.

7. Generative AI

Video Session (Page 388)

Do it yourself.

Ai Reboot (Page 394)

1. labeled, unlabeled 2. conditional 3. Text generation 4. Voice synthesis

Ai Task (Page 397)

Do it yourself.

Exercise

Unsolved Questions

SECTION A (Objective Type Questions)

Ai Quiz

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

B. 1. Generative Adversarial Network 2. Translation 3. In education
4. film production 5. LLM 6. Musick.ai 7. Chatbots
8. Weak AI 9. plagiarism, copyright 10. DeepFaceLab, FaceSwap

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

SECTION B (Subjective Type Questions)

A. 1. Generative AI refers to AI systems that can generate new content such as text, images, music, or code by learning patterns from existing data. It mimics human-like creativity in producing original outputs.

2. GANs are used for creating realistic images, style transfer and data-augmentation.

3. The primary objective is to model the joint probability of input and output data to generate new data that resembles the original dataset.

4. Text generation is the process by which AI models produce human-like written content based on learned patterns from large text datasets. Text generation involves producing human-like written content using models like GPT, which predict the next word in a sequence based on learned patterns from large text datasets.
5. Audio generation uses AI to create realistic speech, sound effects, or music by analyzing and learning from audio patterns, as seen in tools like ElevenLabs and Musick.ai.
6. LLMs may produce biased or factually incorrect content and lack understanding of context or emotions behind the data they process.
7. Ensure transparency, prevent misuse, avoid biased training data, and always verify outputs to avoid misinformation or ethical misuse.
8. Deepfakes are AI-generated fake videos or images that manipulate faces or voices. They pose a threat by spreading misinformation and harming reputations.
9. Chatbots provide instant responses, 24/7 service, and accurate information, reducing wait times and improving customer experience in banking.
10. An API key is a unique code (a string of letters and numbers) that allows a program to access a specific service on the internet. It acts like a password or a digital key that lets your program communicate with an external system. For example, an API key lets a program connect to and use a service, like Google's Gemini AI.

B. Long answer type questions.

1.

Generative AI	Discriminative AI
Creates new data similar to original; detects anomalies	Classifies or predicts labels; separates classes
Learns full data distribution; generates new data	Learns decision boundary; distinguishes classes
Understands data structure to generate new instances	Learns function to predict class labels accurately
Can use labelled or unlabelled data	Requires labelled data
Used in image/music generation, data augmentation, anomaly detection	Used in email filtering, image recognition, sentiment analysis
Examples: GANs, VAEs, GPT, Naive Bayes (generative use)	Examples: SVM, Logistic Regression, Random Forests, Decision Trees

2. Image generation involves creating new visuals based on patterns learned from existing datasets. AI models analyse the features of input images and generate new images that exhibit similar traits. AI models identify and replicate visual features like shapes, colours, textures, and structures.

Some applications of image generation are as follows:

- Graphic design: AI tools assist designers by generating unique images or enhancing existing ones.
- Art creation: AI-generated art is used for digital artwork, abstract designs, and even traditional portrait styles.
- Fashion and product design: AI can generate new clothing designs or create virtual prototypes for product development.
- Advertising and marketing: Brands use AI to create targeted ads or promotional content that resonates with different audiences.
- Film production: AI-generated scenes, backgrounds, and visual effects are used in movies, reducing production costs and time.

Some examples of image generation tools are Canva, DALL-E, Copilot Stability AI & Stable Diffusion, Runway, Artbreeder, etc.

3. Some applications of audio generation are as follows:
   - Music Composition: AI generates original music tracks in various genres, emulating the style of famous musicians or creating entirely new soundscapes.
   - Voice Synthesis: AI can generate realistic speech, imitating specific voices, accents, or even creating entirely new ones.
   - Sound Design: AI can generate sound effects for movies, games, or applications by learning from existing audio libraries.
   - Podcasts and Audiobooks: AI-generated speech can create narrations or podcasts that sound natural, even for long-form content.
   - Speech-to-Text and Text-to-Speech: AI models transcribe spoken words into text and also generate natural-sounding speech from written text.

   Some examples of audio generation are Meta AI's Voicebox, Google's Music LM, Amper Music, Descript's Overdub, Aiva Technologies, ElevenLabs, etc.
4. A Large Language Model (LLM) is an advanced deep learning architecture specifically designed for a wide range of Natural Language Processing (NLP) tasks. These tasks include but are not limited to text generation, classification, question answering in a conversational format, and translation between languages. The term "large" refers to the vast scale of data and parameters these models are trained on, often involving extensive datasets consisting of text, code, and other forms of written content. A Large Language Model (LLM) is an AI model trained on massive text datasets to understand and generate human-like language. It uses transformer-based architecture to predict and produce meaningful text. Examples include GPT and BERT, which are used in chatbots, summarizers, and content creation tools.

5. Popular LLMs include the following:
    - OpenAI's GPT-4o: Multimodal AI processing text, images, and audio; faster and cheaper than GPT-4 Turbo; great for chatbots and automation.
    - Google's Gemini 2.0 Flash: Processes text, images, and speech; integrates Google services; ideal for search and content generation.
    - Meta's LLaMA 3.1: Open-source, efficient model used in research and language tasks.
    - Anthropic's Claude 3.5: Focuses on safe, transparent AI for customer support and advisory.
    - Mistral AI's Mixtral 8x7B: Cost-effective, high-performance model for edge computing and mobile apps.
    - Cohere's Command R+: Scalable AI for enterprise data processing and automated decisions.
    - Amazon's Titan: Enhances business and cloud services within Amazon's ecosystem.
    - Microsoft's Phi-3: Lightweight, cost-effective AI for IoT, embedded, and mobile devices.

    These models support chatbots, content creation, research, and automation across industries.
6. Some risks associated with LLMs are as follows:
    - Bias in responses: LLMs may reflect social and cultural biases from their training data, leading to unfair or offensive outputs.
    - Privacy risks: They might accidentally reveal sensitive or personal information from training data, especially in sectors like healthcare and finance.
    - Susceptibility to manipulation: Malicious users can exploit LLMs with harmful prompts to generate inappropriate or misleading content.
    - Lack of transparency: LLMs don't explain how responses are generated or cite sources, making fact-checking difficult.
    - Potential for misuse: LLMs can be used unethically to create deepfakes, spread disinformation, or automate scams.
7. The quality of the output largely depends on how we frame these prompts. Here are few key points to keep in mind while framing prompts:
    - Be specific: Provide detailed instructions for clearer responses.
    - Add context: Help AI understand the background or situation.
    - Set desired format: Specify how you want the answer (e.g., bullet points).
    - Mention the audience: Indicate who the answer is for to get the right level.
    - Use personas: Ask AI to adopt roles (e.g., teacher, doctor) for expert-like replies.
    - Refine prompts: Adjust and improve your prompts if the response isn't right.
8. Generative AI raises several ethical and social concerns as given below:
    - Deepfake technology creates realistic fake videos/images, causing misinformation and privacy issues.

- Bias and discrimination in AI can lead to unfair treatment in facial recognition and hiring.
- Plagiarism and copyright concerns arise when AI-generated content is used without proper credit.
- Lack of transparency can reduce trust; users must disclose AI assistance to avoid dishonesty.
- Ethical use requires:
  - Being honest about AI-generated content
  - Respecting copyright laws
  - Following institutional guidelines on AI use

C. Competency-based/Application-based questions.

1. a. WanderWorld can use Generative Adversarial Networks (GANs) for creating realistic destination images and Large Language Models (LLMs) for generating personalised travel itineraries and textual content.
   b. Generative AI can power chatbots to respond with natural, human-like language, provide instant travel suggestions, booking help, and destination facts, making the customer experience more interactive and engaging.
   c. A challenge may be the risk of spreading inaccurate or biased information, especially when promoting eco-sensitive destinations that require factual accuracy.
   d. Generative AI can quickly create high-quality, photorealistic visuals tailored to the target audience's preferences, helping attract attention and improve marketing impact.
   e. The agency can ensure accuracy by fact-checking outputs, using verified datasets, and incorporating human oversight in the content creation and approval process.
2. GANs can manipulate visuals and generate highly personalised content that might distort users' perceptions, reinforce biases, or promote unrealistic standards. Ethical concerns include user manipulation, consent, privacy issues, and the potential misuse of deepfakes or synthetic media for misinformation.
3. The suitable AI model is a Large Language Model (LLM) or Generative AI model trained for enterprise analytics, capable of processing vast transaction data, extracting patterns, identifying trends, and supporting automated decisions.
4. Generative AI can help create visually stunning images of fashion products, generate background scenes, and simulate fashion styles. This helps in designing eye-catching visuals for websites and social media platforms, enhancing brand appeal and customer engagement.

Assertion and Reasoning Questions.

5. a.
6. c.

Statement–based questions

7. c.
8. a.

Ai In Life (Page 424)

Do it yourself.

Ai Deep Thinking (Page 425)

Do it yourself.

Ai Lab (Page 425)

Do it yourself.

Ai Ready (Page 426)

Do it yourself.

8. Data Storytelling

Video Session (Page 429)

Do it yourself.

Ai Task (Page 431)

Do it yourself.

Ai Reboot (Page 432)

a. Storytelling is the art of using a sequence of events to communicate a message or meaning in a clear, engaging, and relatable way. It helps make information more memorable and emotionally connected for the audience.

b. The plot is the sequence of events that make up the story. The conflict is the central problem or challenge that drives the plot forward and keeps the audience interested.

Ai Task (Page 434)

Do it yourself.

Video Session (Page 437)

Do it yourself.

Video Session (Page 439)

Do it yourself.

Ai Task (Page 441)

Do it yourself.

Video Session (Page 451)

Do it yourself.

Ai Task (Page 451)

Do it yourself.

Exercise

Unsolved Questions

SECTION A (Objective Type Questions)

Ai Quiz

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

B. 1. Folklore 2. numbers, figures 3. Narrative 4. denouement
5. Data visualization 6. wordcloud 7. Numeric data
8. Pie Chart 9. GPS 10. Ethics

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

SECTION B (Subjective Type Questions)

A. 1. The essential elements of an effective data story are data, narrative, and visuals. Data provides the evidence, narrative gives meaning and context, and visuals help simplify and present information clearly.

2. Narrative connects the data with the audience by adding context, emotion, and meaning. It helps explain why the data matters and what action should be taken based on the insights.

3. The five key stages are:
   - Exposition
   - Rising Action
   - Climax
   - Falling Action
   - Resolution

4. In data storytelling, Freytag's Pyramid serves as a useful method for analysing literature and films, allowing readers to break down a story's progression and understand its deeper meaning. By applying this structure, writers can ensure their stories maintain a natural flow, and readers can gain deeper insights into the mechanics of storytelling.
5. Text data refers to data that exists in the form of written words, such as tweets, customer reviews, or comments. It is unstructured and often analyzed to derive patterns and insights.
6. Candlestick charts are a type of data visualization used to show price movements in financial markets. Each candlestick shows opening, closing, high, and low prices in a given time frame.
7. Steps include:
   - Understand your audience
   - Gather and analyze data
   - Create a narrative
   - Use visuals to support the story
   - Share the story with clarity and purpose
8. Two ethical principles are as follows:
   - Accuracy: Ensure data is factual and not misleading.
   - Transparency: Clearly explain sources and methodology behind the data presented.

B. 1. Data often exists in its raw form as numbers and figures. When data is visually represented—using tools like bar graphs, pie charts, or scatter plots—it becomes part of Data Visualization, which helps simplify and communicate complex information. For instance, a bar chart could show monthly sales performance, or a heatmap could reveal areas with the highest internet usage in a city. Visualization not only simplifies complex information but also helps highlight patterns and relationships making data easier to understand and act upon.

2. In today's data-driven world, companies and organisations rely on data storytelling to effectively communicate insights. By combining engaging narratives with visuals, they make complex data accessible and memorable for diverse audiences. For instance, a company might use data storytelling to illustrate how customer feedback has led to product innovations, helping stakeholders understand the value of listening to their audience. Similarly, governments use data stories to explain public health trends, like vaccination coverage, in ways that citizens can understand.

   In today's data-driven world, data storytelling is becoming an essential tool across industries. Organisations and brands increasingly leverage it as a powerful method to communicate insights effectively, strengthen their messaging, and build client trust. By transforming raw data into structured narratives, data storytelling simplifies complex information, making it more accessible and easier to understand for diverse audiences, including stakeholders, decision-makers, and customers.

3. Data Storytelling is powerful because of following reasons:
   - **Makes insights memorable:** Data storytelling transforms complex insights into structured narratives, enhancing audience comprehension and retention, which leads to informed decision-making.
   - **Persuades stakeholders:** By presenting key insights in a compelling narratives, data storytelling aligns business and technical stakeholders, fostering consensus and strategic action.
   - **Engages the audience:** Combining storytelling with data increases engagement, making analytical findings more relatable, impactful, and easier to interpret.
   - **Bridges the gap between technology and people:** It translates technical data into digestible insights, ensuring that analytics drive real-world applications and organisational change.
   - **Enhances data literacy:** Empowering employees with storytelling techniques improves data literacy, fostering a culture of data-driven decision-making across teams.
   - **Drives action and influence:** By contextualising data within real-world scenarios, data storytelling inspires action, influences decision-makers, and reinforces the relevance of insights. By merging the power of data with the universal appeal of storytelling, this skill is now more essential than ever in turning raw information into actionable strategies and fostering a data-driven culture.
4. The significance of a narrative stems from the fact that it explains what is happening within the data set. It provides context and meaning, as well as relevance and clarity. A narrative directs the audience's attention and keeps them engaged by showing them where to look and what not to miss. Finally,when narrative and visuals are combined, they enlighten an audience. When the right visuals and narrative are combined with the right data, you have a data story that can motivate and influence change.

   Good stories do not arise from data alone; they must be uncovered through data relationships. A closer look reveals how each data point links to the other.
5. The climax is the most intense and dramatic moment in the story. It represents a turning point, often involving a major decision, revelation, or confrontation. This moment is critical because it determines the direction of the rest of the narrative. The climax is often when the protagonist faces their greatest challenge or realises an important truth about themselves or the world around them. It is also the point where emotions reach their peak, whether it be excitement, tension, or sorrow.

   Example: Through advanced analysis, the team discovers that introducing a same-day delivery service in urban areas could significantly boost customer retention (loyalty). This insight shifts the focus towards actionable solutions.

C. Competency-based/Application-based questions.

1. i. The primary insight was that fuel efficiency was lower during peak traffic hours, which led to higher costs and delayed deliveries.

ii. The company proposed rescheduling deliveries to off-peak hours and optimising delivery routes to reduce fuel consumption and improve efficiency.

iii. Optimising routes and delivery schedules reduces fuel costs, saves time, and avoids traffic delays, leading to better delivery performance and improved customer satisfaction.

iv. a.

v. b.

2. c.

3. They conducted a customer feedback survey to gather data on product satisfaction, user experience, feature requests, and usage patterns, and created a data storytelling presentation to share findings.

4. i. Customer Reviews and Feedback → Text Data → Word Cloud

ii. Monthly Sales Figures → Numerical Data → Bar Graph

iii. Store Locations and Regional Performance → Geographical Data → Map

Assertion and Reasoning Questions.

5. a.

6. d.

Statement–based questions

7. c.

8. c.

Ai In Life (Page 463)

Do it yourself.

Ai Deep Thinking (Page 463)

Do it yourself.

Ai Lab (Page 463)

Do it yourself.

Class Activity (Page 465)

Do it yourself.

Ai Ready (Page 466)

Do it yourself.