Artificial Intelligence

Code 417

Class: X

Part B

Unit 1: AI Project Cycle & Ethical Frameworks

A. Short answer type questions.

1. **What is the purpose of data exploration in the AI project cycle?**
   Ans: Data exploration is the process of analysing data to discover patterns and gain insights using data visualisation methods like graphs. It simplifies complex data, helps in selecting AI models and makes it easy to communicate insights to others.

2. **What role does Natural Language Processing (NLP) play in AI?**
   Ans: NLP enables AI to understand, interpret, and generate human language. It is used in applications like chatbots, translation, sentiment analysis, and speech recognition, allowing machines to communicate effectively with humans.

3. **How does rights-based ethics differ from utility-based ethics?**
   Ans: Rights-based ethics focus on individual rights and freedoms, ensuring fairness and respect for personal autonomy. In contrast, utility-based ethics prioritise the greatest good for the greatest number, sometimes at the expense of individual rights.

4. **What is the significance of non-maleficence in healthcare?**
   Ans: Non-maleficence is the ethical principle of "do no harm." In healthcare, it ensures that medical decisions, treatments, and AI applications do not cause unnecessary harm or suffering to patients.

5. **What is bioethics in the context of AI?**
   Ans: Bioethics in AI addresses ethical issues arising from the use of AI in healthcare and life sciences. It includes concerns about patient privacy, fairness in medical decisions, transparency of AI models, and the responsible use of AI in medical research and treatment.

B. Long answer type questions.

1. **What is data visualisation?**
   Ans: Data visualisation is the graphical representation of data using charts, graphs, maps, and other visual formats. It helps in identifying patterns, trends, and relationships in data, making it easier to interpret and analyse. Common tools for data visualisation include bar charts, pie charts, scatter plots, and heatmaps. In AI, data visualisation is essential for understanding datasets, communicating insights, and making data-driven decisions.

2. **How do bioethical principles apply when AI is used in fields like healthcare and life sciences?**
   Ans: Bioethical principles ensure the responsible and fair use of AI in healthcare and life sciences. Key principles include:
   - **Autonomy:** AI should respect patient choices and privacy.

- **Beneficence:** AI should be used to enhance patient care and improve health outcomes.
- **Non-maleficence:** AI should not cause harm, ensuring medical decisions are safe and unbiased.
- **Justice:** AI should be accessible to all, avoiding discrimination in medical treatment.

For example, AI-powered diagnostic tools must be transparent and reliable to prevent biases in medical decisions.

3. **Explain the domains of AI and describe their contributions to various industries with relevant examples**.

Ans: Artificial Intelligence (AI) is a broad field that includes different domains, each contributing to advancements in various industries. Three key domains of AI are Statistical Data, Computer Vision, and Natural Language Processing (NLP).

Statistical Data

This domain involves using AI and machine learning techniques to analyse large datasets, identify patterns, and make predictions. It helps in data-driven decision-making across industries. The contributions are as follows:

Price comparison websites: Price comparison websites collect data from various sellers to help users compare prices of a product in one place.

Search engines: Search engines like Google, Bing, etc., rely heavily on statistical data to deliver relevant search results to users.

Website recommendations: Companies like Amazon, Flipkart, and others use large volumes of statistical data to recommend products based on previous search results of a user. They display products and advertisements based on user data to enhance engagement and overall user experience.

Computer Vision

Computer Vision enables machines to interpret and analyse visual information from images or videos. It is used in facial recognition, object detection, and autonomous systems. The contributions are as follows:

Agriculture monitoring: Computer vision helps in agriculture by analysing crops, soil, and farming conditions using AI-powered image recognition.

Self-driving automobiles: Self-driving automobiles use computer vision extensively. Automated cars from companies like Tesla can detect 360-degree movements of pedestrians, cyclists, vehicles, and other objects. Computer vision helps them detect and analyse objects in real time and take decisions like braking, stopping, or to keep driving.

Surveillance systems: Surveillance systems use computer vision to monitor and analyse real-time video footage for security and safety purposes.

Natural Language Processing (NLP)

NLP enables AI systems to understand, interpret, and generate human language, making it crucial for communication-based applications. The contributions are as follows:

Email filters or spam filters: Gmail and other mailing platforms use email filters or spam filters to separate unwanted or malicious messages, known as spam, from genuine emails. These filters work

by scanning email content for specific keywords or patterns to identify and block spam and ensure that only relevant and safe messages reach the user's inbox.

Sentiment analysis: NLP is used to analyse and determine the sentiment or emotional tone of text data (positive, negative, or neutral), such as social media posts, customer reviews, and news articles.

Virtual assistants: These technologies facilitate smooth interaction between humans and machines, allowing users to ask questions, set reminders, or complete tasks through voice or text commands. For example, Siri, Alexa, Google Assistant.

Text summarisation: NLP can automatically summarise long pieces of text by extracting key information and condensing it into shorter versions.

Machine translation: NLP powers machine translation systems like Google Translate and Microsoft Translator. It enables the automatic translation of text from one language to another, making cross-lingual communication more accessible and efficient.

4. **Suhana works for a company where she has been assigned the task of developing a project using AI project cycle. She knows that the first stage is scoping the problem. Help her list the remaining stages she must go through to complete the project**.

Ans: Suhana has already identified **Problem Scoping** as the first stage of the **AI Project Cycle**. The remaining stages she must go through to complete the project are:

Data Acquisition – Collect relevant and high-quality data needed for the AI model.

Data Exploration – Analyse and visualise the collected data to identify patterns and insights.

Modelling – Train and develop an AI model using suitable algorithms.

Evaluation – Assess the model's performance to ensure accuracy and reliability.

Deployment – Implement the trained model into a real-world system for practical use.

5. **Discuss the importance of an ethical framework for AI. Explain how ethical principles contribute to the responsible development and deployment of AI systems.**

Ans: An ethical framework is crucial for AI to ensure fairness, accountability, and transparency. Without ethical guidelines, AI systems may cause harm, discrimination, or privacy violations. Ethical principles guide AI development in the following ways:

- **Fairness:** Prevents biases and discrimination in AI models.
- **Transparency:** Ensures AI decisions are explainable and accountable.
- **Privacy & Security:** Protects user data and prevents misuse.
- **Responsibility:** Ensures AI systems align with human values and social good.

  For example, ethical AI is essential in hiring processes to avoid bias, and in healthcare, AI-driven diagnostics must be accurate and unbiased to ensure patient safety.

Case -based Questions

1. **A large social media company, SocialNet, collects vast amounts of personal data from its users to target advertisements more effectively. While this data collection improves user experience by personalising content and ads, the company has faced backlash over its handling of sensitive information, including allegations of data breaches and unauthorised sharing of data with third parties. The company's growth has raised questions about whether it is balancing its profit-driven motives with its responsibility to protect user privacy and data security. Examine the ethical challenges associated with SocialNet's data privacy practices, particularly in terms of rights-based ethics. In what ways does the company's approach to data collection possibly violate users' rights, and what steps can be implemented to safeguard individual autonomy and ensure the protection of personal dignity?**

   Ans: Ethical Challenges:

   SocialNet's data collection practices raise ethical concerns related to users' right to privacy, autonomy, and personal dignity. Rights-based ethics emphasizes the inherent rights of individuals, such as:

   - **Right to Privacy:** Users have a fundamental right to control their personal information.
   - **Right to Informed Consent:** Users should have clear and transparent choices about data sharing.
   - **Right to Security:** Companies must protect user data from breaches and unauthorised access.

   Potential Violations:

   - **Lack of Transparency:** SocialNet may not provide clear disclosures on how personal data is used.
   - **Unauthorised Data Sharing:** Selling or sharing data with third parties without consent violates privacy rights.
   - **Data Breaches:** Weak security measures can expose sensitive user information.

   Steps to Safeguard User Autonomy & Data Protection:

   - **Implement Transparent Policies:** Provide users with clear, accessible privacy policies.
   - **Strengthen Security Measures:** Use encryption, multi-factor authentication, and regular audits.
   - **Enable User Control:** Allow users to opt in/out of data collection and customise privacy settings.
   - **Accountability & Oversight:** Establish independent regulatory oversight to ensure ethical compliance.

2. **MediPharm, a global pharmaceutical company, has developed a life-saving drug for a rare and dangerous disease. The company sets an extremely high price for the drug, arguing that the cost of research and development justifies the price. However, critics argue that the high price puts the drug out of reach for many patients who need it, especially in low-income countries. The company is under increasing pressure from the public, healthcare organisations, and**

governments to lower the price and make the drug more accessible to those in need. Evaluate the ethical concerns related to MediPharm's pricing strategy using the principles of beneficence and non-maleficence. How does the company's pricing affect patient well-being, and what ethical guidelines should influence the company in balancing profitability with the responsibility to provide affordable healthcare?

Ans: Ethical Concerns:

MediPharm's pricing strategy conflicts with the ethical principles of:

- **Beneficence (Doing Good):** The company should prioritize patient well-being by making life-saving drugs accessible.
- **Non-Maleficence (Do No Harm):** High prices may prevent patients from accessing essential treatment, leading to suffering or death.

Impact on Patient Well-Being:

- **Limited Access to Treatment:** Many low-income patients may not afford the drug, worsening health inequalities.
- **Social Injustice:** The high price benefits only the wealthy, excluding disadvantaged groups.
- **Public Distrust:** The perception of prioritising profits over lives can damage the company's reputation.

Ethical Guidelines for Balancing Profit & Accessibility:

- **Tiered Pricing Models:** Offer lower prices in low-income regions while maintaining sustainable profits.
- **Collaboration with Governments & NGOs:** Work with public health organisations to subsidise costs.
- **Transparency in Pricing:** Clearly justify the costs associated with drug development.
- **Corporate Social Responsibility (CSR):** Invest in programs that improve global healthcare access.

3. **TalentMatch, a leading recruitment firm, has recently deployed an AI-powered hiring tool to streamline its recruitment process. The tool uses machine learning algorithms to screen resumes, assess candidate qualifications, and rank applicants. However, after implementing the system, the company noticed that candidates from certain demographic groups were being systematically overlooked by the AI, resulting in unintentional discrimination. TalentMatch must now evaluate the ethical implications of its AI system to ensure it aligns with fairness and equality principles. Examine the ethical challenges surrounding TalentMatch's AI recruitment tool through the lens of right-based ethics. How might the system's potential biases undermine fairness, and what measures should be implemented to ensure the AI tool offers equal opportunities to candidates from all backgrounds?**

Ans: Ethical Challenges:

The AI hiring tool raises concerns about **fairness, equality, and non-discrimination**, violating rights-based ethics principles such as:

- **Right to Fair Treatment:** Every candidate deserves an equal opportunity.
- **Right to Non-Discrimination:** AI biases should not disadvantage individuals based on race, gender, or background.
- **Right to Transparency:** Applicants should understand how AI decisions are made.

How AI Biases Undermine Fairness:

- **Algorithmic Discrimination:** If the AI model is trained on biased historical data, it may favour certain demographics.
- **Lack of Explainability:** Candidates may not understand why they were rejected.
- **Reinforcement of Existing Inequities:** If the AI prioritises past hiring trends, it may exclude underrepresented groups.

Measures to Ensure Fair AI Recruitment:

- **Bias Audits & Regular Monitoring:** Continuously test AI for discriminatory patterns.
- **Diverse Training Data:** Ensure datasets represent diverse demographics.
- **Human Oversight:** Use AI as an assistive tool rather than an autonomous decision-maker.
- **Transparency & Accountability:** Allow candidates to challenge unfair decisions and improve model explainability.

Unit 2: AI Advanced Concepts of Modelling in AI

A. Short answer type questions.

1. What is machine learning?

Ans: Machine Learning is a field of AI that enables machines to learn on their own and improve with time through experience. In ML, machines learn from data fed to them during the training phase and use this knowledge to improve their performance in making accurate predictions.

2. Differentiate between classification and regression model.

Ans:

Features	**Classification**	**Regression**
Definition	Predicts discrete labels or categories.	Predicts continuous numerical values.
Output Type	Categorical (e.g., "spam" or "not spam").	Continuous (e.g., predicting house prices).
Examples	Email spam detection, fraud detection.	Stock price prediction, temperature forecasting.
Algorithms Used	Decision Trees, Random Forest, SVM, Neural Networks.	Linear Regression, Polynomial Regression, Neural Networks.

3. **What are ANNs?**
   Ans: Artificial Neural Networks (ANNs) are computing systems inspired by the human brain that process data using layers of interconnected neurons. ANNs are widely used in machine learning for tasks like image recognition, speech processing, and pattern detection.

4. **What happens in the output layer of a neural network?**
   Ans: The output layer produces the final prediction or classification result of the neural network. It transforms processed data into meaningful outputs using activation functions.

5. **Define CNN.**
   Ans: Convolutional Neural Networks (CNNs) are a type of deep learning model designed for processing structured grid-like data, such as images. CNNs use convolutional layers to detect patterns like edges, textures, and objects, making them effective for image recognition and computer vision tasks.

6. **Define association in unsupervised learning.**
   Ans: Association in unsupervised learning refers to discovering relationships between variables in large datasets. It identifies patterns and correlations, commonly used in market basket analysis to find product purchase associations (e.g., "Customers who buy bread often buy butter").

B. Long answer type questions.

1. **Differentiate between the rule-based approach and the learning-based approach.**
   Ans:

Rule-based approach	**Learning-based approach**
The machine follows the rules defined by the developer.	The machine learns on its own from the data.
AI is achieved through rule-based technique.	AI is achieved through learning technique.
It typically uses labelled data.	It can handle both labelled and unlabelled data.
It may require less training time.	It requires more training time.

2. **What are supervised, unsupervised and reinforcement learning? Explain with examples?**
   Ans: Supervised learning is a machine learning technique where an algorithm learns from labelled data. Each training example consists of an **input** and a corresponding **correct output (label)**. The model analyses these labelled examples, recognises patterns, and applies this knowledge to make predictions on new, unseen data. For example, email spam detection—the model is trained on emails labelled as "spam" or "not spam."
   Unsupervised learning works on **unlabelled** datasets. This means that the data that is fed to the machine is random, and there is a possibility that the person who is training the model might not have any information regarding it. Unsupervised learning models are used to identify relationships,

patterns, and trends in the data that is fed into them. For example, customer segmentation—grouping customers based on purchasing behaviour without predefined labels.

Reinforcement learning is a machine learning approach where an intelligent agent, i.e., a computer program, interacts with the environment in such a way that it can gain maximum rewards. This AI model learns through trial and error by testing different actions and selecting those that yield the highest rewards, without being explicitly programmed for the task. For example, A robot learning to walk — it receives rewards for successful movements and improves over time through trial and error.

3. **Explain neural networks. Also, describe the role of input, hidden, and output layers in a neural network.**

Ans: A neural network is based on the model of the human brain and consists of multiple layers. Each layer performs a specific task and passes the results to the next layer. The different layers of the neural networks are listed below:

Input layer: This is the first layer of the neural network. The job of an input layer is to receive data and pass it on to the rest of the network. No processing occurs at the input layer.

Hidden layers: Hidden layers are the processing units of a neural network, positioned between the input and output layers. The number of hidden layers and the nodes within them depend on the complexity of the network's function. Each node in a hidden layer applies a machine learning algorithm to process the data received from the previous layer.

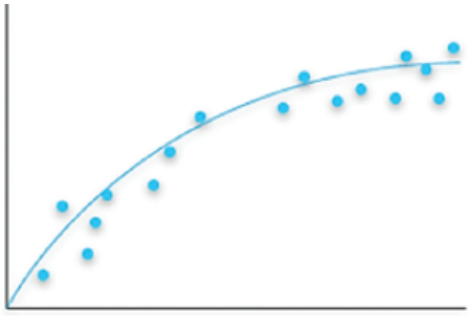
Output layer: The output layer receives the data from the last hidden layer and fault tolerant to the user. Similar to the input layer, the output layer also does not process data. It serves as a user interface, presenting the final outcome of the network's computations to the user.

4. **Describe the following AI models and draw an appropriate diagram:**

a. Regression

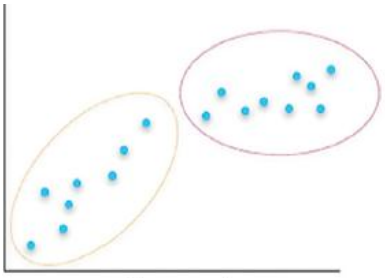
b. Clustering

Ans: Regression: Regression is another type of supervised learning model used to predict continuous numerical values, i.e., it works on continuous data. Continuous data means data that can have any value within a certain range. For example, you might want to predict the future price of a stock. Stock prices are continuous variables (because stock prices can be any number within a range), and regression models can be used to make predictions based on historical stock data and other relevant factors. For instance, we can use an employee's past salary data to train the AI model to predict their future salary. The algorithm learns from this data and creates a solid line on a graph that represents the pattern in the salaries. In the given graph, the blue dots represent the data values we have, i.e., the past salaries, and the solid line represents the function that helps us predict future values.



Regression

Clustering: Clustering is an unsupervised machine learning technique where the model groups a dataset into different clusters or groups based on similarities in data points. Unlike supervised learning, clustering does not require labelled data—the model identifies patterns or trends on its own to categorise data points. The data in this model is usually unlabelled or random. In the graph given here, the algorithm has identified two groups or clusters based on the patterns in the data. The datasets within each oval are more similar to each other compared to those in the other oval.


Clustering

5. **Neural networks are said to be modelled the way how neurons in the human brain behave. A similar system is mimicked by the AI machine to perform certain tasks. Explain how neural networks work in an AI model and mention any three features of neural networks.**
   Ans: Neural Networks are computational networks that are at the heart of deep learning algorithms, a subfield of Artificial Intelligence. Similar to how our brains learn from experiences, neural networks learn from examples to understand new situations. A neural network is initially trained on large amounts of input data. The network recognises the patterns in this data, learns from it using machine learning techniques and can then make predictions on a new dataset. It is a fast and efficient way to solve problems for which the dataset is very large, such as in images and videos.
   Features
   - Artificial neural networks are extremely powerful computational algorithms or models.
   - Neural Network systems are modeled on the structure and function of the human brain and nervous system.
   - The most powerful feature of neural networks is that once trained, they can independently process new data, take decisions and make predictions without human intervention.

Case -based Questions

1. **Imagine that you have to analyse social media posts to identify patterns in user sentiment and group similar posts together. The dataset consists of text data from social media posts (unlabelled), and your goal is to group posts with similar sentiment. Which AI model will you use?**
   Ans: Since the dataset consists of unlabelled text data, the best AI model would be clustering using unsupervised learning.
   Model Used: Clustering
   Reason: It groups similar posts together without predefined labels.
   Example: Grouping social media posts by sentiment (positive, neutral, negative).
2. **A fruit exporting company had a sorting plant that sorted the two types of labelled fruits, i.e., apples and oranges. A classification AI model was being used to sort the fruits. Now, the company has decided to export the fruits in larger quantities where the variety of fruits could be different every time. As an AI developer, how will you modify your algorithm to adjust to this change?**

Ans: The company's previous classification model worked only for apples and oranges (labelled data). Since new varieties of fruits may vary, a more scalable approach is required.
Modified Approach: Use an unsupervised learning clustering model (e.g., K-Means) to group similar fruits based on features like size, colour, and texture.
Hybrid Approach: First, use clustering to group unknown fruits, then retrain a classification model to recognise the new categories.

3. **Identify the type of learning (supervised, unsupervised, reinforcement learning) on which the following case studies most likely based. [CBSE]**
   a. **A company wants to predict customer churn based on past purchasing behaviour, demographics, and customer interactions. They have a dataset with labelled examples of customers who churned and those who did not.**
   b. **A social media platform wants to group users based on their interests and behaviour to recommend relevant content. They have a large dataset of user interactions but no predefined categories.**
   c. **An autonomous vehicle is learning to navigate through a city environment. It receives feedback in the form of rewards for reaching its destination safely and penalties for traffic violations.**
   d. **A healthcare provider wants to identify patterns in patient data to personalise treatment plans. They have a dataset with various patient attributes but no predefined labels indicating specific treatment plans.**
   e. **A manufacturing company wants to optimise its production process by detecting anomalies in sensor data from machinery. They have a dataset with examples of normal and anomalous behaviour.**

   Which type of learning is this case study most likely based on?
   Ans:

Case Study	**Type of Learning**	**Reason**
a. Predicting customer churn (with labelled data).	Supervised Learning	Labelled data indicates which customers churned.
b. Grouping users based on behaviour (no predefined labels).	Unsupervised Learning	No predefined categories; patterns are discovered.
c. Autonomous vehicle learning from rewards/penalties.	Reinforcement Learning	AI learns through rewards and penalties.
d. Identifying patient patterns for treatment personalization (no predefined labels).	Unsupervised Learning	Patterns are discovered in patient data.
e. Detecting anomalies in sensor data (normal vs. anomalous).	Supervised Learning	Labelled dataset with examples of normal/anomalous behaviour.

4. **Identify the type of model (classification, regression, clustering, association model) on which the following case studies most likely based. [CBSE]**
   a. **A bank wants to predict whether a loan applicant will "default" or "non-default" on their loan payments. They have a dataset containing information such as income, credit score, loan amount, and employment status.**
   b. **A real estate agency wants to predict the selling price of houses based on various features such as size, location, and the number of bedrooms and bathrooms. They have a dataset containing historical sales data.**
   c. **A marketing company wants to segment its customer base into distinct groups based on purchasing behaviour for targeted marketing campaigns. They have a dataset containing information such as purchase history, frequency of purchases, and amount spent.**
   d. **A grocery store wants to identify associations between different products purchased by customers to understand which products are commonly bought together. They have a transaction dataset containing records of items purchased together during each transaction.**

   Ans:

Case Study	**Type of Model**	**Reason**
a. Predicting loan default vs. non-default.	Classification	Predicting categories (default or non-default).
b. Predicting house prices based on features.	Regression	Predicting a continuous value (price).
c. Segmenting customers based on behaviour.	Clustering	Finding patterns in unlabelled data.
d. Identifying frequently bought products together.	Association	Finding relationships between purchased items.

5. **A healthcare provider wants to improve patient care by predicting the length of hospital stays for different medical conditions. They have a dataset containing patient demographics, medical history, and treatment details. The task involves: [CBSE]**
   a. predicting whether a patient will have a short or long hospital stay.
   b. predicting the number of days a patient will stay in the hospital.
   c. segmenting patients into groups with similar characteristics for personalised treatment plans.
   d. identifying patterns in patient treatments and outcomes.

   Identify the type of model (classification, regression, clustering, and association model) in the above given tasks.

Ans:

Task	Type of Mode	Reason
a. Predicting short vs. long hospital stay.	Classification	Predicting discrete labels (short/long stay).
b. Predicting the number of hospital days.	Regression	Predicting a continuous value (days).
c. Grouping patients with similar characteristics.	Clustering	Identifying patterns in unlabelled data.
d. Identifying patterns in treatments and outcomes.	Association	Discovering hidden relationships.

6. **Convert the following scenarios to perceptron: [CBSE]**

Context 1: A manager is deciding whether to approve a work-from-home request from an employee.

Factors:

- **Does the employee perform well when working remotely?**
- **Are there any upcoming team meetings or collaborative projects?**
- **Does the company's policy support remote work?**
- **Is it beneficial for both the employee and the company?**

Context 2: A homeowner is deciding whether to invest in solar panels for their house or not.

Factors:

- **Do I have a sufficient average amount of sunlight in my area?**
- **Are there any available incentives or rebates for installing solar panels?**
- **Does install solar panels impact the value of my home?**
- **Does solar energy lead to environmental benefits?**

Ans: **Context 1: Work-from-Home Request Decision**

Factors	Input (1 = Yes, 0 = No)
Performs well remotely?	1
Upcoming team meetings?	0
Company policy supports remote work?	1
Environmental benefits?	1
Perceptron Output	Approved (1) or Denied (0)

Context 2: Solar Panel Investment Decision

Factors	Input (1 = Yes, 0 = No)
Sufficient sunlight?	1
Incentives available?	1
Impact home value positively?	1
Beneficial for both employee & company?	1
Perceptron Output	Invest (1) or Not Invest (0)

Unit 3: Evaluating Models

A. Short answer type questions.

1. **Define Recall and list the types of cases it considers.**
   Ans: Recall can be defined as the fraction of positive cases that are correctly identified, i.e., True Positives to the sum of True Positives and False Negatives (instances incorrectly identified as negative). It is also known as True Positive Rate or Sensitivity. Recall considers True Positive and False Negative cases.

2. **Which metric calculates the percentage of true positive cases versus all the cases where the prediction is true?**
   Ans: Precision calculates the percentage of true positive cases versus all the cases where the prediction is true.

3. **What is classification accuracy, and is it always suitable for evaluating AI models?**
   Ans: Classification accuracy is defined as the correct predictions out of all observations. In other words, accuracy tells us how often the model's predictions match with the actual outcomes.
   Limitations:
   - **Not suitable** for **imbalanced datasets** (e.g., detecting rare diseases).
   - A model with high accuracy may still perform poorly if false positives or false negatives are critical.

4. **Which metrics does the F1 score take into consideration?**
   Ans: The **F1 Score** is the harmonic mean of **Precision** and **Recall**, balancing false positives and false negatives:
   $$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. **What is meant by the terms "True Positive" and "True Negative" in model evaluation?**
   Ans: True Positive in model evaluation is a case where predictions and reality match and prediction is Positive.
   True Negative in model evaluation is a case where predictions and reality match and prediction is Negative.

6. **What are the four potential outcomes of confusion matrix?**
   Ans:

   True Positive (TP): Correctly predicted positive cases.
   False Positive (FP): Incorrectly predicted positive cases (Type I error).
   True Negative (TN): Correctly predicted negative cases.
   False Negative (FN): Incorrectly predicted negative cases (Type II error).

7. **Why should we avoid using training data for evaluation?**
   Ans:
   Overfitting Risk: The model may memorize patterns from training data but fail on unseen data.
   Lack of Generalization: The model's real-world performance is unknown without testing on separate data.
   Best Practice: Use **train-test split** or **cross-validation** to properly evaluate performance.

B. Long answer type questions.

1. **Give one real-world example each where the cost of false negative and false positive can be high.**
   Ans: Situation where cost of False Negative is high: In case of autonomous cars if a model gives a False Negative while detecting pedestrians or other vehicles, it could result in accidents and result in loss of life.
   Situations where cost of False Positive is high: A False positive by an airport security screening model that predicts a security threat, when it predicts a threat when there isn't any, can lead to flight delays, inconvenience, and missed flights for travellers. The cost to the airline is also high.

2. **What is the importance of the confusion matrix?**
   Ans: A confusion matrix is a summarised table used to analyse and assess the performance of an AI model. The matrix compares the actual target values with those predicted by the model. This allows us to visualise how well our classification model is performing and what kinds of errors it is making.

3. **Automated trade industry has developed an AI model, which predicts the selling and purchasing of automobiles. During testing, the AI model came up with the following predictions:**

The Confusion Matrix	Reality: 1	Reality: 0
Predicted: 1	55 (TP)	12 (FP)
Predicted: 0	10 (FN)	20 (TN)

a. How many total tests have been performed in the above scenario?

b. Calculate the Accuracy, Precision, Recall, and F1 Score.

Ans:

- Accuracy: (TP + TN) / Total tests] * 100 = [75 / 97] * 100 = 77.32%
- Precision: [TP / (TP + FP)] * 100 = [55 / 67] * 100 = 0.8209 * 100 OR 82.09%
- Recall: [TP / (TP + FN)] * 100 = [55 / 65] * 100 = 0.8462 * 100 OR 84.62%
- F1 Score: [2 * (Precision * Recall) / (Precision + Recall)] * 100 = [2 * (0.8209 * 0.8462) / (0.8209 + 0.8462)] * 100 ≈ 0.83

Summary:

- Total tests performed: 97
- Accuracy: 77.32%
- Precision: 82.09%
- Recall: 84.62%
- F1 Score: 0. 83

4. **Taking examples, discuss the importance of the F1 score.**

   Ans: In order to assess if the performance of a model is good, we need two measures: Recall and Precision. In some cases, you may have a high Precision but low Recall and in others, low Precision but high Recall. But since both the measures are important, there is a need of a metric which takes both Precision and Recall into account. The metric that takes into account both these parameters is F1 Score. F1 score can be defined as the measure of balance between Precision and Recall. F1 score combines both Precision and Recall into a single number to give a better overall picture of how well the model is performing.

5. **Recently the country was shaken up by a series of earthquakes that caused a huge damage to the people as well as the infrastructure. To address this issue, an AI model has been created to predict whether there is a chance of earthquake or not. The confusion matrix for the same is also given.**

Confusion Matrix		Reality	
		Yes	No
Prediction	Yes	50	05
	No	25	20

   a. How many total cases are True Negative in the scenario?

   Ans: 20

   b. Calculate Precision, Recall, and F1 score.

   Ans:

   Precision: [TP / (TP + FP)] * 100

   = [50 / (50 + 5)] * 100

   = [50 / 55] * 100

   ≈ 0.9091 * 100 OR 90.91%

   Recall (Sensitivity): TP / (TP + FN)

= [50 / (50 + 25)] * 100
= [50 / 75] * 100
≈ 0.6667 + 100 OR 66.67%
F1 Score: 2 * (Precision * Recall) / (Precision + Recall)
= [2 * (0.9091 * 0.6667) / (0.9091 + 0.6667)] * 100
≈ 0.77

6. **List the ethical concerns around model evaluation.**

Ans: The ethical concerns around model evaluation are:

Bias

Bias in model evaluation refers to systematic errors that lead to unfair outcomes, often disadvantaging specific groups based on factors like race, gender, or socio-economic status.

Concerns include:

Data bias: Models trained on biased data can reflect and even amplify those biases in their predictions, leading to discriminatory outcomes. For example, underrepresentation of certain groups in the training dataset can result in poorer model performance for those groups.

Algorithmic bias: The choice of algorithms or evaluation metrics may prioritise certain outcomes over others, favouring one group while disadvantaging another.

Impact: Biased models can increase inequalities and reduce trust in AI systems, especially in sensitive applications like hiring, lending, or healthcare.

Transparency: Transparency refers to the clarity and openness of the evaluation process, including the criteria, metrics, and methods used to assess the model's performance. This involves an honest explanation of how the chosen evaluation metrics work and produce results, without keeping any information hidden. Concerns include:

Opaque metrics: If stakeholders do not understand how evaluation metrics are calculated, it can weaken the trust in the model's fairness and utility.

Hidden assumptions: Evaluation processes might include unspoken assumptions or decisions that impact the outcomes, such as selecting metrics that prioritise accuracy over fairness.

Impact: Lack of transparency can make it difficult to identify flaws, hold developers accountable, and ensure the model meets ethical standards.

Accountability: Accountability refers to taking responsibility for the effects and outcomes of a model’s evaluation process. It involves:

Clear ownership: Identifying who is responsible for the evaluation metrics and any resulting decisions.

Remediation processes: Having a plan to correct or mitigate harm if the model’s predictions negatively impact individuals or groups.

Stakeholder engagement: Involving those affected by the model’s decisions, ensuring their concerns are heard and addressed.

7. **Explain train-test split method.**
   Ans: Train-test split is a commonly used method for evaluating the performance of a machine learning algorithm. This technique works with any supervised learning algorithm: classification and regression. The process involves partitioning the dataset into two subsets. The first subset, known as the **training dataset**, is used to train the model. The second subset, the **testing dataset**, remains unseen during training. It is used to evaluate the model's performance by providing input data to the model, generating predictions, and comparing them against the actual outcomes. The goal of train-test split is to evaluate how well a machine learning model will perform on unseen data, simulating its practical use. In real-world scenarios, the model is trained on existing data with known inputs and outputs, and then used to predict outcomes for new data where the target values are unknown.

Case -based Questions

1. **You are developing an AI system to classify customer reviews as positive or negative for a restaurant review website, where the restaurant owners are more interested in good reviews given by customers. Which evaluation metric out of Recall and Precision would be more important when assessing the system's performance and why?**
   Ans: Choosing Between Precision and Recall for Restaurant Review Classification
   - **More Important Metric: Precision**
   - **Reason:** Since restaurant owners are more interested in **good reviews**, they want to minimize the number of **false positives** (incorrectly classifying a negative review as positive).
   - **Formula:**

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

   - High **Precision** ensures that only genuinely positive reviews are labelled as such, avoiding misleading classifications.

2. **Imagine you are a coach for a district level cricket team for schools, and you are thinking of implementing an AI-based player evaluation system to assess the performance of your players. The system will analyse different parameters, such as batting performance, bowling performance, fitness, and fielding. Discuss both the cost of false positives (wrongly identifying players with lower skills as top performers) and false negatives (overlooking talented players) for this scenario.**
   Ans:
   False Positive (Wrongly Identifying a Weak Player as a Top Performer)
   - **Impact:** Weak players might be selected for key matches.
   - **Consequence:** The team's overall performance may suffer, leading to losses.
   - **Ethical Issue:** Deserving players may miss out on opportunities.

False Negative (Overlooking Talented Players)

- **Impact:** Skilled players might not be selected.
- **Consequence:** The team may lose potential star performers.
- **Ethical Issue:** Players may become discouraged, impacting their growth.

Conclusion: False negatives are more costly in this scenario, as they can result in missing out on potential talent, which is critical for long-term success.

3. **A water quality monitoring system classifies water samples as either contaminated (1) or safe (0). Out of 1000 water samples:**
   True Positives (TP): 160 samples were correctly identified as contaminated.
   False Positives (FP): 30 samples were incorrectly flagged as contaminated.
   True Negatives (TN): 780 samples were correctly identified as safe.
   False Negatives (FN): 30 samples were incorrectly labelled as safe.
   Construct the confusion matrix and compute the performance metrics: accuracy, precision, recall, and F1-score.
   Ans: Confusion Matrix:

	Predicted Contaminated (1)	Predicted Safe (0)
Actual Contaminated (1)	TP = 160	FN = 30
Actual Safe (0)	FP = 30	TN = 780

Performance Matrix:

Accuracy = (TP + TN)/(TP + TN + FP + FN)
= (160 +780)/ (160 + 780 + 30 + 30)
= 940/1000 = 0.94

Precision = TP/(TP + FP)
= 160/(160 + 30)
= 160/190 = 0.842

Recall = TP/(TP + FN)
= 160/ (160 +30)
= 160/190 = 0.842

F1 Score = 2 x [(Precision x Recall)/(Precision + recall)]
= 2 x [(0.842 x 0.842)/ (0.842 + 0.842)]
= 2 x [0.7089/1.684]
= 2 x 0.4209 = 0.8419

4. **A traffic monitoring model predicts whether a driver will commit traffic violation (1) or not (0) within a year. Out of 1000 drivers:**
   True Positives (TP): 80 drivers were correctly predicted to commit violation.
   False Positives (FP): 20 drivers were incorrectly predicted to commit violation.
   True Negatives (TN): 880 drivers were correctly predicted not to commit violation.
   False Negatives (FN): 20 drivers who committed violation were not predicted.
   Construct the confusion matrix and compute the performance metrics: accuracy, precision, recall, and F1-score.
   Ans:

Confusion Matrix:

	Predicted Violation (1)	Predicted No Violation (0)
Actual Violation (1)	TP = 80	FN = 20
Actual No Violation (0)	FP = 20	TN = 880

[Find the performance matrix as shown in Answer 3]

Unit 4: Statistical Data

A. Short answer type questions.

1. **What is no-code AI, and what are the benefits of using no-code AI tools?**
   Ans: No-Code AI refers to AI tools and platforms that allow users to build, train, and deploy AI models without writing code. These platforms use drag-and-drop interfaces, visual workflows, and pre-built templates to simplify AI development.

2. **Define low-code AI.**
   Ans: Low-Code AI refers to AI development that requires minimal coding by combining pre-built components with some scripting or customization to fine-tune the model. It offers more flexibility than No-Code AI while still being simpler than traditional AI programming.

3. **What types of tasks can be accomplished with no-code AI tools?**
   Ans:
   Text Analysis – Sentiment analysis, text classification, chatbot development.
   Image Processing – Object detection, facial recognition, image classification.
   Data Analysis – Predictive analytics, anomaly detection, data visualization.
   Automation – Automated decision-making, workflow automation, document processing.

4. **Name any two popular web-based no-code AI tools.**
   Ans:
   Teachable Machine – A Google tool for training machine learning models without coding.
   Microsoft Lobe – A simple no-code AI tool for image classification.

5. **List three statistical concepts important for data analysis.**
   Ans: Three Important Statistical Concepts for Data Analysis
   Mean (Average) – The sum of all values divided by the number of values.
   Median – The middle value when data is sorted in ascending order.
   Standard Deviation – Measures the dispersion of data points from the mean.

6. **What is Orange, and how is it used in data science?**
   Ans: Orange is an open-source data visualization and analysis tool that allows users to perform machine learning, data mining, and statistical analysis through a graphical interface. It is widely used in education, research, and industry for tasks such as clustering, classification, and data exploration.

7. **Mention the primary tabs available in Orange.**
   Ans:
   Data – For loading and manipulating datasets.
   Visualize – For creating graphs and visualizations.
   Model – For machine learning model training.
   Evaluate – For model performance evaluation.
   Unsupervised – For clustering and association analysis.

8. **What are some widgets in Orange for statistical analysis?**
   Ans:
   Distributions – Displays histograms and distributions of variables.
   Box Plot – Shows the spread and outliers in a dataset.
   Statistics – Computes basic statistics like mean, median, and standard deviation.
   Scatter Plot – Helps visualize relationships between two numerical variables.

9. **Calculate the mean, mode, and median for the following dataset: 2, 3, 4, 6, 7, 8, 3, 4, 5, 6, 7, 8, 9, 3**

Ans:

$$\text{Mean} = \frac{\sum \text{Values}}{\text{Number of Values}}$$

$$= \frac{2+3+4+6+7+8+3+4+5+6+7+8+9+3}{14} = \frac{71}{14} = 5.07$$

Mode (Most Frequent Value): 3 (appears **three times** in the dataset).

Median (Middle Value in Sorted Data):

Sorted Data: **2, 3, 3, 3, 4, 4, 5, 6, 6, 7, 7, 8, 8, 9**

Since there are **14 values**, the median is the **average of the 7th and 8th values**:

$$\frac{5+6}{2} = 5.5$$

10. **Write the steps to plot bar chart using Orange for the following dataset: 4, 5, 6, 7, 8, 9, 12, 34, 23, 34.**

Ans: Steps:

1. **Load Orange** and open the **Canvas workspace**.
2. **Drag and drop the File widget** and upload the dataset.
3. **Connect the File widget to the "Distributions" widget** for a bar chart representation.
4. **Select the feature (column) containing the data** from the dataset.
5. **Adjust bin size and appearance** to visualize data distribution.
6. **Click Apply** to generate the bar chart.

B. Long answer type questions.

1. **Explain the key differences between no-code and low-code AI platforms with examples.**

Ans:

No-code AI platforms allow users to create AI models **without any coding skills**. They use simple drag-and-drop tools. For example, Google Teachable Machine lets you train models using images, sounds, or poses easily.

Low-code AI platforms need **a small amount of coding** and give more control to the user. They are suitable for users who know a little programming. For example, Microsoft Power Platform allows you to build smart apps with minimal code.

2. **Describe the role of statistical analysis in artificial intelligence and data science.**

Ans: Statistical analysis plays an important role in artificial intelligence (AI) and data science by helping to **collect, organize, and interpret data**. It helps AI systems to understand patterns, relationships, and trends in data, which is necessary for making predictions and decisions.

For example, by using statistics, an AI model can learn to predict the weather or identify objects in images. Statistical tools also help measure how accurate and reliable AI models are, making sure they work well.

3. **Compare and contrast different widgets in Orange used for data transformation and visualisation.**
   Ans: In Orange Data Mining, widgets are categorized into different groups based on their functionality, such as Data, Visualize, Transform, and Model. Here's a comparison of different widgets used for Data Transformation and Visualisation:
   - **Data Transformation widgets** modify or prepare data for analysis, while **Visualization widgets** represent data graphically.
   - Some widgets (e.g., **Impute**, **Normalize**) improve data quality, whereas others (e.g., **Scatter Plot**, **Heat Map**) help in pattern recognition.
   - **Feature Constructor** enables the creation of new attributes, whereas **Box Plot** helps in analyzing existing data distributions.

4. **Explain some applications of Data Science.**
   Ans: Some applications of Data Science are stated below.
   Fraud and Risk Detection
   Data scientists studied this data to understand customer behaviour and predict the risk of loan defaults. This helped banks decide who should get loans and how much, reducing losses. They also used data to offer products based on customers' purchasing power.
   Genetics and Genomics
   Data science is used in genetics and genomics to study how our genes (DNA) affect our health. By analysing genetic information, doctors can offer personalised treatments based on how genes influence diseases and responses to medicines.
   Search Engines
   Search engines like Google, Bing, Ask, and AOL extensively use data science algorithms to provide users with the best search results based on their search key in a fraction of seconds.
   Targeted Advertising
   Data science is extensively used in targeted digital marketing. Display banners on websites as well as billboards in airports use data science algorithms to analyse user behaviour and target advertisements based on it.
   Website Recommendations
   Popular websites and streaming platforms such as Amazon Prime, X, Google Play, Netflix, LinkedIn, and
   IMDB use data science to recommend products, movies, and shows based on the user's previous buying patterns and search history.
   Airline Route Planning

Airline route planning can greatly benefit from data science by helping to improve profitability. Data science techniques can predict flight delays, analyse which flight routes are in demand, and assist in determining whether it is more cost-effective to take direct routes or make halts in between.

Case -based Questions

1. **As the CEO of a small e-commerce startup, you are eager to leverage artificial intelligence to enhance your platform's user experience and drive sales. However, your team lacks the technical expertise to develop and deploy AI-powered solutions. What would be your recommendations for the CEO? [CBSE]**

   Ans:

   - Leverage No-Code/Low-Code AI Platforms
   - Partner with AI-as-a-Service Providers
   - Upskill the Team in AI Basics
   - Collaborate with AI Consultants or Freelancers
   - Focus on AI-Driven Personalization & Marketing

2. **Neha, a healthcare analyst, wants to study patient recovery times but cannot analyse the entire dataset. She takes a sample and applies statistical methods. Explain statistical sampling, how descriptive statistics (mean, median, mode) help summarise data, and how distributions, probability, and variance influence her analysis in healthcare decision-making.**

   Ans: A. Statistical Sampling in Healthcare

   Neha uses statistical sampling to analyse patient recovery times without examining the entire dataset. Common methods include random sampling, stratified sampling, and systematic sampling to ensure representative data.

   Descriptive statistics help summarise data:

   - **Mean** (average recovery time)
   - **Median** (central value, useful with outliers)
   - **Mode** (most frequent recovery time)

   Distributions, probability, and variance influence analysis:

   - **Probability** predicts recovery likelihood.
   - **Distributions** (e.g., normal or skewed) reveal recovery trends.
   - **Variance & standard deviation** measure spread, helping in risk assessment.

Unit 5: Computer Vision

A. Short answer type questions.

1. **What do you mean by resolution?**

   Ans: The number of pixels in an image is called its **resolution**. It determines the level of detail or clarity of images displayed on screens or captured by cameras.

2. **Define CNN.**

   Ans: A Convolutional Neural Network is a deep Learning algorithm that is commonly used in image recognition and processing. CNN analyses an image, extracts the best features and reduces its size to make it manageable, while still preserving its important features. This helps it to differentiate one image from the other.

3. **Enlist two smartphone apps that utilise Computer Vision technology. How have these apps improved your efficiency or convenience in daily tasks?**

   Ans: Google Lens

   Purpose: Google Lens uses Computer Vision to identify objects, text, landmarks, and products through your phone's camera.

   Snapchat

   Purpose: Snapchat uses Computer Vision for features like Lenses and Filters, which recognize your face and apply virtual effects in real-time. This enhances user experience by allowing fun and interactive photo and video editing.

4. **What are considered good features of an image during image processing?**

   Ans: The features like corners are easy to find as their exact location can be pinpointed in the image. Thus, corners are always good features to extract from an image followed by the edges.

5. **What are pixels?**

   Ans: The word 'pixel' stands for picture element. Every digital photograph is made up of tiny elements called pixels. A pixel is the smallest unit of information that constitutes texts, images, or videos on a computer.

6. **How is an RGB image different from a grayscale image?**

   Ans: An RGB image has three colour channels (Red, Green, Blue) and represents full-colour visuals. Each pixel stores three values for colour mixing. A grayscale image has only one channel, storing intensity values (black to white). RGB is used for colour photos and displays, while grayscale is common in black-and-white photography and medical imaging.

7. **Briefly describe the purpose of the convolution operator in image processing.**
   Ans: The convolution operator in image processing helps detect patterns like edges, textures, and shapes by applying a filter (kernel) to an image. It slides over the image, performing mathematical operations to highlight important features. This process is key in image recognition, sharpening, and blurring in Computer Vision and CNNs.

8. **Determine the colour of a pixel based on its RGB values mentioned below:**
   a. R=0, B=0, G=0 **b. R=255, B=255, G=255**
   c. R=0, B=0, G=255 **d. R=0, B=255, G=0**
   Ans: Here is the colour corresponding to each set of RGB values:
   a. **R=0, G=0, B=0**
      Colour: Black (no colour, all channels are 0).
   b. **R=255, G=255, B=255**
      Colour: White (all channels are at their maximum value).
   c. **R=0, G=0, B=255**
      Colour: Blue (only the blue channel is at maximum value).
   d. **R=0, G=255, B=0**
      Colour: Green (only the green channel is at maximum value).

B. Long answer type questions.

1. **How do computers store RGB images?**
   Ans: Every RGB image is stored on a computer in the form of three different channels - the Red channel, the Green channel and the Blue channel. Each channel contains a number of pixels, with the value of each pixel ranging from 0 to 255. When all channels are combined together, they form a coloured image. This means that in an RGB image, each pixel has a set of three different values which together give colour to that particular pixel.

2. **Explain CV and provide any two applications of it.**
   Ans: The Computer Vision domain of artificial intelligence enables machines to interpret visual data, process it and analyse it using algorithms and methods to interpret real-world phenomena. It helps machines to derive meaningful information from digital images, videos and other visual inputs and take actions based on that information.
   Applications:
   Face filters: This is one of the popular applications used in apps like Instagram and Snapchat. A face filter is a filter applied to photographs, or videos in real time, to make the face look more attractive. You can also use it to combine a face with animal features to give it a funny appearance.
   Facial Recognition: With smart homes becoming popular, computer vision is being used for making homes more secure. Computer Vision facial recognition is used to verify the identity of the visitors and guests and to maintain a log of the visitors. This technology is also used in social networking applications for detecting faces and tagging friends.

3. **What is meant by the pixel value of a digital image?**
   Ans: Each pixel in a digital image on a computer has a pixel value which determines its brightness or colour. The most common pixel format is the byte image, where this value is stored as an 8-bit integer having a range of possible values from 0 to 255. Typically, zero is considered as no colour or black and 255 is considered to be full colour or white.

4. **Describe the CV tasks for a single object in an image.**
   Ans: The CV tasks for a single object in an image:
   Image Classification: This task involves assigning a label to the entire image based on its contents.
   Image Classification plus Localisation: This is the task which involves both processes of identifying what object is present in the image and at the same time identifying at what location that object is present in that image.

5. **State the difference between CV and human vision.**
   Ans:

Feature	**Computer Vision**	**Human Vision**
Processing	Uses algorithms and machine learning to analyse images.	Uses the brain and sensory perception to interpret visuals.
Adaptability	Requires training on datasets to recognise new objects.	Can quickly recognize and adapt to new objects with prior experience.
Speed of Learning	Requires large datasets and training.	Learns from experience with minimal exposure.
Understanding Context	Struggles with abstract reasoning and context.	Easily understands context, emotions, and intent.

6. **The Face Lock feature of a smart phone is an example of Computer Vision. Briefly discuss this feature.**
   Ans: The Face Lock feature in smartphones uses Computer Vision to scan and recognize a user's face for unlocking the device. It captures facial features like eye position and nose shape, then compares them with stored data. If the features match, access is granted; otherwise, it stays locked. Some advanced systems use AI and infrared sensors for improved accuracy, even in low light. 3D facial recognition adds extra security, making it harder to bypass with photos or videos. This feature provides both convenience and security for smartphone users.

7. **What is the difference between 'classification' and 'classification and localisation' in context with Computer Vision tasks? Give an example to illustrate this difference.**

   Ans: Classification:

   - Identifies what object is present in an image.
   - Does not specify where the object is located.
   - Example: A model detects an image contains a cat but does not indicate its position.

   Classification & Localisation:

   - Identifies what object is present and where it is in the image.
   - Provides bounding box coordinates around the detected object.
   - Example: The model detects a cat in an image and draws a box around it to indicate its location.

   Illustration Example:

   1. Classification: "This image contains a dog."
   2. Classification & Localisation: "This image contains a dog at (x, y) with a bounding box."

8. **Explain the role and function of the convolution operator in CNNs.**

   Ans: The convolution operator is a fundamental mathematical operation used in image processing. It involves multiplying two arrays of numbers, element-wise, to produce a third array. In image processing, convolution is a common tool used for image editing to apply filters or effects, such as blurring, sharpening, outlining or embossing on an image. An image convolution is simply an element-wise multiplication of the image array with another array called the kernel followed by sum. This results in forming a convolution matrix or filtered image.

9. **What are the different layers in Convolutional Neural Network? What features are likely to be detected by the initial layers of a neural network and how is it different from what is detected by the later layers? [CBSE]**

   Ans: A Convolutional Neural Network (CNN) has multiple layers that process images step by step. The initial layers detect basic features like edges, lines, and colours. As the image moves deeper into the network, the later layers recognize complex patterns, such as shapes, objects, and faces. The Convolutional Layer extracts features, the Pooling Layer reduces data size, and the Fully Connected Layer makes final decisions. Early layers focus on small details, while deeper layers understand the whole object.

10. **Apart from text-based search, search engines also carry out image-based searches. How does Computer Vision help in this task?**

    Ans: Visual search algorithms in search engines use computer vision technology to help you search for different objects using real world images. CV compares different features of the input image to its database of images, analyses the image features and gives us the search result. Computer vision, combined with machine learning allows the device not only to see the image, but also to interpret what is in the picture, helping make decisions based on it.

11. How is Computer Vision used in the following fields?

a. Healthcare b. Warehouses

Ans:

a. **Healthcare**: Medical imaging has greatly benefited from computer vision. It not only creates and analyses images, but also acts as an assistant and helps doctors to better understand a patient's health condition. CV analyses X-Rays, CT scans and MRI and is used to read and convert 2D scanned images into interactive 3D models. This results in increase in the accuracy and efficiency of diagnosis since the machines can identify details invisible to the human eye.

b. **Warehouses**: CV can be used in warehouses to remove human error during the receiving and storing process of products by automating the scanning and data entry process for inventory management. Robots equipped with CV accurately pick parcels and pack them. CV also automatically checks the order against the contents.

12. "Computer Vision is the basis of autonomous vehicles." Justify.

Ans: Autonomous driving involves identifying objects, getting navigational routes and monitoring of surroundings. Automated cars from companies like Tesla can detect the 360-degree movements of pedestrians, vehicles, road signs and traffic lights and create 3D maps. CV helps them detect and analyse objects in real-time and take decisions like breaking, stopping or keep driving.

13. What are the two ways to describe pixel count?

Ans: The two ways to describe pixel count are:

Width by height: When referring to resolution as 'Width by Height', it means describing the number of pixels in an image. For example, a monitor with a resolution of 1280×1024 means there are 1280 pixels horizontally (from side to side) and 1024 pixels vertically (from top to bottom).

Single number: Resolution can also be expressed as a single number. For example, if we say that a camera is a 5-megapixel camera (a megapixel is a million pixels), it means that the number of pixels along the width multiplied by the number of pixels along the height of the image taken by the camera equals 5 million pixels. In the above-given example of the 1280×1024 monitor, the resolution could also be expressed as:

1280 × 1024 = 1,310,720, or 1.31 megapixels.

14. State the tasks involved in instance segmentation.

Ans: This task finds pixels that belong to the object in the image. It involves accurately identifying and labelling individual instances of objects within the image and categorising and labelling them. To achieve this, a segmentation algorithm takes an image as input and generates a set of regions or segments, where each pixel is labelled according to the specific object instance it belongs to.

Case -based Questions

1. **Imagine you are developing a smart home security system. What Computer Vision features could enhance the system's ability to detect and alert the owners of the home about security threats?**
   Ans: A smart home security system can use Computer Vision for:
   - Facial Recognition – Identifies known residents and alerts for unknown visitors.
   - Motion Detection – Detects unusual movements and triggers alerts.
   - Object Recognition – Identifies suspicious objects like weapons or unattended bags.
   - Intruder Detection – Recognizes forced entry or unauthorized access.
   - Night Vision & Thermal Imaging – Enhances detection in low-light conditions.
   - License Plate Recognition – Monitors and records vehicles near the home.

2. **Think of a scenario where you are developing a medical imaging application. How can Computer Vision assist in detecting and diagnosing diseases from X-rays, MRIs, or CT scans?**
   Ans: Computer Vision in medical imaging helps detect diseases by analysing X-rays, MRIs, and CT scans for abnormalities like tumours or fractures. It enhances image clarity, detects patterns, and assists doctors in early diagnosis, improving accuracy and patient care.

3. **Imagine you are building an AI project for identifying different monuments and buildings from photographs. The task involves extracting features from images in order to identify objects for further processing. Which library or module can you use to achieve this? Justify.**
   Ans: You can use OpenCV and TensorFlow/Keras for identifying monuments from photographs.
   - OpenCV helps in feature extraction using techniques like SIFT, SURF, and ORB, which detect key points and patterns in images.
   - TensorFlow/Keras allows training a deep learning model (CNNs like ResNet or VGG) for accurate monument classification.
     These libraries are widely used for image recognition, offering efficient feature extraction, scalability, and real-time processing.

4. **Imagine you have a smartphone camera app that can recognise objects. When you point your camera at a dog, the app identifies it as a dog, analysing patterns and features in the image. Behind the scenes, the app's software processes the image, detecting edges, shapes, and colours, then compares these features to a vast database to make accurate identifications. Identify the technology used in the above scenario and explain the way it works. [CBSE]**
   Ans: The technology used is Computer Vision with Deep Learning. The app processes the image using CNNs, detects edges, shapes, and colours, and compares these features with a pre-trained model to identify objects like a dog. This enables real-time and accurate object recognition.

5. **Imagine you are a researcher tasked with improving workplace safety in a manufacturing environment. You decide to employ Computer Vision technology to enhance safety measures. [CBSE]**

   Ans: I would use Computer Vision to enhance workplace safety by implementing real-time monitoring for detecting hazards like unsafe worker behaviour, improper PPE usage, and machine malfunctions. AI-powered cameras can identify risks, send alerts, and automate compliance checks, reducing accidents and ensuring a safer manufacturing environment.

6. "**Agriculture is an industry where precision and efficiency are crucial for sustainable production. Traditional farming methods often rely on manual labour and visual inspection, which can be time-consuming and error-prone. However, advancements in Computer Vision technology offer promising solutions to optimise various agricultural processes. Agricultural drones equipped with high-resolution cameras and Computer Vision algorithms are increasingly being used to monitor crop health, detect diseases, and assess crop yields." Answer the following questions based on the case study mentioned above:**
   a. **How does the integration of Computer Vision technology with drones improve efficiency agricultural practices compared to traditional methods?**
   b. **What are some key indicators or parameters that Computer Vision algorithms can analyse to assess crop health and detect diseases? [CBSE]**

   Ans:

   a. **Improved Efficiency with Computer Vision Drones**
      Computer Vision-equipped drones improve efficiency by automating crop monitoring, reducing the need for manual inspections. They cover large fields quickly, providing real-time data on crop health and detecting issues early, leading to faster decision-making and improved yields.
   b. **Key Indicators for Crop Health & Disease Detection**
      Computer Vision algorithms analyse factors like leaf colour changes (chlorosis), plant height, wilting, pest infestations, and nutrient deficiencies. They also use NDVI (Normalised Difference Vegetation Index) and thermal imaging to assess plant stress and disease outbreaks accurately.

7. **You are tasked with developing a Computer Vision system for a self-driving car company. The system needs to accurately detect and classify various objects on the road to ensure safe navigation. Imagine you are working on improving the object detection algorithm for the self-driving car's Computer Vision system. During testing, you notice that the system occasionally misclassifies pedestrians as cyclists, especially in low-light conditions. How would you approach addressing this issue? What steps would you take to enhance the accuracy of pedestrian detection while ensuring the system's overall performance and reliability on the road?**
   Ans: To improve pedestrian detection in low-light conditions, I would collect more nighttime data, use infrared or thermal imaging, and apply data augmentation (e.g., brightness adjustments).

Enhancing the model with better feature extraction and multi-scale detection can improve accuracy. Continuous testing and fine-tuning would ensure reliable performance on the road.

Unit 6: Natural Language Processing

A. Short answer type questions.

1. **Define NLP.**
   Ans: Natural Language Processing is a field of artificial intelligence that enables computers to understand and interpret human (natural) language. NLP takes a verbal or written input, processes it and analyses it, based on which appropriate action can be taken.

2. **How do companies use NLP to get feedback from customers regarding their products and services?**
   Ans: Companies use Natural Language Processing applications, such as sentiment analysis, to identify the emotions in the text and to categorise opinion about their products and services as 'good', 'bad' or 'neutral'. This process can be used to identify emotions in text even when it is not clearly expressed and enables companies to understand what customers think about their brand and image. It helps not only to understand what people like or dislike but understand what affects a customer's choice in deciding what to buy.

3. **Name some popular virtual assistants that use NLP to help us in our daily lives.**
   Ans: Some popular virtual assistants are Google Assistant, Copilot and Siri.

4. **List the common applications of script bots.**
   Ans: Script bots are used for simple functions like answering frequently asked questions, setting appointments and on messaging apps to give predefined responses.

5. **Give an example of sentences using a word with the same spelling but different meanings.**
   Ans: "The bat is hanging upside down on the tree"
   "Anju bought a new bat for the cricket match finale"

6. **Give the stem and lemma of the word 'studies'.**
   Ans: Stem: studi. Lemma: study.

7. **What does the word "bag" in the "Bag of Words" algorithm symbolise?**
   Ans: The name "bag" symbolises that the algorithm is not concerned with where the words occur in the corpus i.e. the sequence of tokens, but aims at getting unique words from the corpus and the frequency of their occurrence.

8. **List the steps involved in the "BoW" algorithm.**
   Ans:
   Step 1: Text Normalisation - Collect data and pre-process it.
   Step 2: Create Dictionary - Make a list of all the unique words occurring in the corpus. (Vocabulary).
   Step 3: Create document vectors for each document - Find out how many times the unique words from the document have occurred.
   Step 4: Create document vectors for all the documents.

B. Long answer type questions.

1. **How does the human brain process sound?**
   Ans: Our brain keeps processing the sounds that it hears and tries to make sense out of them. Sound travels through air, enters the ear and reaches the eardrum through the ear canal. The sound striking the eardrum is converted into a neuron impulse and gets transported to the brain. This signal is then processed by the brain to derive its meaning and helps us give the required response.

2. **How does automatic summarization help us make sense of a large amount of textual data?**
   Ans: Automatic summarization helps us make sense of large amounts of textual data by condensing key information into a shorter, coherent version while retaining the most important points. It allows us to:
   - **Save Time** – Instead of reading lengthy documents, reports, or articles, a summary provides a quick understanding of the main ideas.
   - **Improve Comprehension** – Summarization highlights essential concepts, making complex information easier to grasp.
   - **Enhance Decision-Making** – Professionals can make informed decisions based on concise insights extracted from large datasets.
   - **Enable Efficient Searching** – Summaries help users quickly determine whether a document is relevant to their needs.
   - **Support Information Overload Management** – With the vast amount of digital text available, summarization tools help filter and prioritize important content.

3. **What is meant by "perfect syntax, no meaning" in the context of a language? Illustrate with an example?**
   Ans: Sometimes, a sentence can have a correct syntax but it does not mean anything. For example, “Purple elephants dance gracefully on my ceiling.”
   This statement is correct grammatically but does not make any sense.

4. **How does text normalisation help in processing text?**
   Ans: Text normalization helps process text by standardising it, removing inconsistencies like case variations, punctuation, and extra spaces. It ensures uniformity in spelling, numbers, and date

formats, making text easier for machines to analyse. By eliminating noise and handling variations, it improves the accuracy of tasks like sentiment analysis and machine learning.

5. **Describe the following steps involved in text normalisation.**
   a. Sentence Segmentation b. Tokenization
   Ans:
   a. **Sentence Segmentation**: In sentence segmentation, the entire corpus is divided into sentences. Based on punctuation marks the entire corpus is split into sentences.

   b. **Tokenization:** After segmenting the sentences, each sentence is further divided into tokens. Tokenization is the process of separating a piece of text into smaller units called tokens. Token is a term used for any word or number or special character occurring in a sentence. Under tokenisation, every word, number and special character is considered as a separate unit or token.

6. **State the difference between stemming and lemmatization. Give examples to illustrate your answer.**
   Ans: In Stemming, the words left in the corpus are reduced to their root words. Stemming is the process in which the affixes of words are removed and the words are converted to their base form or "stem". Stemming does not take into account if the stemmed word is meaningful or not. It just removes the affixes; hence it is faster. For example, the words – 'programmer, programming and programs' are reduced to 'program' which is meaningful, but 'universal' and 'beautiful' are reduced to 'univers' and 'beauti' respectively after removal of the affix and are not meaningful.
   Lemmatisation too has a similar function, removal of affixes. But the difference is that in lemmatization, the word we get after affix removal, known as lemma, is a meaningful one. Lemmatization understands the context in which the word is used and makes sure that lemma is a word with meaning. Hence it takes a longer time to execute than stemming. For example: 'universal' and 'beautiful' are reduced to 'universe' and 'beauty' respectively after removal of the affix and are meaningful.

7. **Explain how the BoW algorithm creates a document vector using an example.**
   Ans: Let us understand the steps involved in implementing a BoW by taking an example of three documents with one sentence each.
   Document 1: *Hema is learning about AI*
   Document 2: *Hema asked the smart robot KiBo about AI*
   Document 3: *KiBo explained the basic concepts*
   Step 1: **Text Normalisation** - Collecting data and pre-processing it.
   Document 1: [*hema, is, learning, about. ai*]
   Document 2: [*hema, asked, the, smart, robot, kibo, about, ai*]
   Document 3: [k*ibo, explained, the, basic, concepts*]

No tokens have been removed in the stopwords removal step because we have very little data and since the frequency of all the words is almost the same, no word can be said to have lesser value than the other.

Step 2: **Create Dictionary** - Make a list of all the unique words occurring in the corpus. (Vocabulary)
Listing the unique words from all three documents:

hema	is	learning	about	ai	asked	the
smart	robot	kibo	explained	basic	concepts	

Step 3: **Create document vector**

In this step, a table with frequency of unique words in each document is created. The vocabulary i.e. unique words are written in the top row of the table. For each document, in case the word exists, the number of times the word occurs is written in the rows below. If the word does not occur in that document, a 0 is put under it.

For example, for the first document:

hema	is	learning	about	ai	asked	the	smart	robot	kibo	explained	basic	concepts
1	1	1	1	1	0	0	0	0	0	0	0	0

Step 4: Create document vectors for all documents

hema	is	learning	about	ai	asked	the	smart	robot	kibo	explained	basic	concepts
1	1	1	1	1	0	0	0	0	0	0	0	0
1	0	0	1	1	1	1	1	1	1	0	0	0
0	0	0	0	0	0	1	0	0	1	1	1	1

In this table, the header row contains the vocabulary of the corpus and three rows below it corresponds to the three different documents

8. **"In text processing, we pay special attention to the frequency of words occurring in the text." Elaborate.**

Ans: In text processing we pay special attention to the frequency of words occurring in the text, since it gives us valuable insights into the content of the document. Based on the frequency of words that occur in the graph, we can see three categories of words. The words that have the highest occurrence across all the documents of the corpus are considered to have negligible value. These words, termed as **stop words**, do not add much meaning to the text and are usually removed at the pre-processing stage. The words that have moderate occurrence in the corpus are called **frequent words**. These words are valuable since they relate to subject or topic of the documents and occur in sufficient number throughout the documents. The less common words are termed as **rare words**. These words appear the least frequently but contribute greatly to the corpus' meaning. When processing text, we only take frequent and rare words into consideration.

9. Samiksha, a student of class X was exploring the Natural Language Processing domain. She got stuck while performing the text normalisation. Help her to normalise the text on the segmented sentences given below: [CBSE Exam]

Document 1: Akash and Ajay are best friends.
Document 2: Akash likes to play football but Ajay prefers to play online games.
Ans:
Normalization Steps Applied:

1. **Lowercasing** – All text is converted to lowercase to maintain consistency.
2. **Removing Punctuation** – Periods (.) are removed to ensure uniform tokenization.
3. **Tokenization (if needed)** – The text can be split into words for further processing.
4. **Lemmatization/Stemming (if needed)** – Since no word variations exist here, this step is optional.

Normalised Text:

- **Document 1:** akash and ajay are best friends
- **Document 2:** akash likes to play football but ajay prefers to play online games

10. Through a step-by-step process, calculate TF-IDF for the given corpus: [CBSE Exam]

Document 1: Johny Johny Yes Papa, Document 2: Eating sugar? No Papa
Document 3: Telling lies? No Papa Document 4: Open your mouth, Ha! Ha! Ha!

Ans: Step 1: Tokenization (Removing Punctuation & Lowercasing)
We preprocess the text by removing punctuation and converting all words to lowercase.
Processed Documents:

1. Document 1: johny johny yes papa
2. Document 2: eating sugar no papa
3. Document 3: telling lies no papa
4. Document 4: open your mouth ha ha ha

Step 2: TF is calculated for each term in a document.

Term	Doc 1 (TF)	Doc 2 (TF)	Doc 3 (TF)	Doc 4 (TF)
johny	2/4 = 0.5	0	0	0
yes	1/4 = 0.25	0	0	0
papa	1/4 = 0.25	1/4 = 0.25	1/4 = 0.25	0

Term	Doc 1 (TF)	Doc 2 (TF)	Doc 3 (TF)	Doc 4 (TF)
eating	0	1/3 = 0.33	0	0
sugar	0	1/3 = 0.33	0	0
no	0	1/3 = 0.33	1/3 = 0.33	0
telling	0	0	1/3 = 0.33	0
lies	0	0	1/3 = 0.33	0
open	0	0	0	1/4 = 0.25
your	0	0	0	1/4 = 0.25
mouth	0	0	0	1/4 = 0.25
ha	0	0	0	3/4 = 0.75

Step 3: Compute Inverse Document Frequency (IDF) IDF is calculated as:

$$IDF(w) = \log\left(\frac{N}{df(w)}\right)$$

Where:

- N = 4 (Total number of documents)
- df(w) is the number of documents containing the term.

Term	df(w)	IDF(w) = log(4/df(w))
johny	**1**	**log(4/1) = 1.39**
yes	**1**	**log(4/1) = 1.39**
papa	**3**	**log(4/3) = 0.29**
eating	**1**	**log(4/1) = 1.39**
sugar	**1**	**log(4/1) = 1.39**

Term	df(w)	IDF(w) = log(4/df(w))
no	2	log(4/2) = 0.69
telling	1	log(4/1) = 1.39
lies	1	log(4/1) = 1.39
open	1	log(4/1) = 1.39
your	1	log(4/1) = 1.39
mouth	1	log(4/1) = 1.39
ha	1	log(4/1) = 1.39

Step 4: Compute TF-IDF. Multiplying the TF values by their corresponding IDF values:

Term	Doc 1 (TF-IDF)	Doc 2 (TF-IDF)	Doc 3 (TF-IDF)	Doc 4 (TF-IDF)
johny	0.5 × 1.39 = 0.70	0	0	0
yes	0.25 × 1.39 = 0.35	0	0	0
papa	0.25 × 0.29 = 0.07	0.25 × 0.29 = 0.07	0.25 × 0.29 = 0.07	0
eating	0	0.33 × 1.39 = 0.46	0	0
sugar	0	0.33 × 1.39 = 0.46	0	0
no	0	0.33 × 0.69 = 0.23	0.33 × 0.69 = 0.23	0
telling	0	0	0.33 × 1.39 = 0.46	0
lies	0	0	0.33 × 1.39 = 0.46	0
open	0	0	0	0.25 × 1.39 = 0.35

Term	Doc 1 (TF-IDF)	Doc 2 (TF-IDF)	Doc 3 (TF-IDF)	Doc 4 (TF-IDF)
your	0	0	0	0.25 × 1.39 = 0.35
mouth	0	0	0	0.25 × 1.39 = 0.35
ha	0	0	0	0.75 × 1.39 = 1.04

Conclusion

- "Johny" has the highest importance in Document 1.
- "Eating" and "sugar" are most important in Document 2.
- "Telling" and "lies" are most significant in Document 3.
- "Ha" has the highest TF-IDF in Document 4 because it appears three times.

11. With reference to data processing, expand the term 'TF-IDF'. Also, give any two applications of TF-IDF. [CBSE Exam]

Ans: TF-IDF stands for **Term Frequency-Inverse Document Frequency**. It is a statistical measure used in data processing and Natural Language Processing (NLP) to evaluate how important a word is in a document relative to a collection (corpus).
Any two applications of TF-IDF are as follows:
Topic modelling: It helps in predicting the topic for a corpus.

Text summarization and keyword extraction: This can be used to help summarise articles more efficiently or to even determine keywords for a document.

12. Create a document vector table from the following documents by implementing all the four steps of Bag of words model. Also, depict the outcome of each. [CBSE Exam]

Document 1: Neha and Soniya are classmates.
Document 2: Neha likes dancing but Soniya loves to study mathematics.
Ans: **Step 1**: Text Normalization
Document 1: [neha, and, soniya, are, classmates]
Document 2: [neha, likes, dancing, but, soniya, loves, to, study, mathematics]

Step 2: Create Dictionary (Vocabulary) Unique words from all documents:
[neha, and, soniya, are, classmates, likes, dancing, but, loves, to, study, mathematics]

Step 3: Create Document Vector for Document 1

neha	and	soniya	are	classmates	likes	dancing	but	loves	to	study	Mathematics
1	1	1	1	1	0	0	0	0	0	0	0

Step 4: Create Document Vector all documents: 1 and 2

neha	and	soniya	are	classmates	likes	dancing	but	loves	to	study	Mathematics
1	1	1	1	1	0	0	0	0	0	0	0
1	0	1	0	0	1	1	1	1	1	1	1

13. What are stopwords? Why are they removed during text pre-processing?

Ans: Stopwords are common words that appear frequently in a language but do not carry significant meaning in text analysis. Examples include **"is," "the," "and," "in," "on," "a," "to," "with,"** etc. These words are generally not useful for tasks like text classification or sentiment analysis.

Removal of stopwords during text pre-processing

- **Reduces Noise in Text Data** – Stopwords do not contribute meaningful information and can clutter the analysis. Removing them helps focus on important words.
- **Improves Computational Efficiency** – Processing fewer words reduces memory and computation time, making NLP models faster.
- **Enhances Text Mining Accuracy** – By eliminating redundant words, algorithms like TF-IDF and Bag of Words (BoW) produce more meaningful results.

14. How does text classification help us get information easily and efficiently?

Ans: Text classification in NLP can be used to automatically classify or predict a category to which a text belongs without human intervention. Text classification groups documents into predefined categories based on the content and organises it in a way that you find easy to get the information you need. For example, email services use text classification for spam filtering by identifying the contents of each email automatically.

15. Define chatbots. What are its types?

Ans: A chatbot is one of the most popular NLP applications. Chatbots, sometimes known as 'Chat Robots', are user-friendly agents that can converse with humans in natural language. while also carrying out tasks like scheduling appointments, sending reminders, and responding to questions on websites and messaging applications. Chatbots first identify the meaning of the question asked by the user, collect all the information needed to respond to it, and then provide the proper response. As you interact with chatbots, you realise that some of them are traditional chatbots or scripted bots while others are AI-powered and have more capabilities. Based on this, chatbots are broadly divided into two categories, namely script bots and smart bots.

16. **What is the outcome provided by the Bag of Words (BoW) algorithm?**

Ans: The Bag of Words (BoW) algorithm converts a collection of text documents into a numerical representation by creating a document-term matrix (DTM).

Key Outcomes:

- Document-Term Matrix (DTM):
  - Each row represents a document.
  - Each column represents a unique word (feature).
  - The values indicate the frequency of words in each document.
- Text Representation as Vectors:
  - Each document is transformed into a vector of word counts, making it suitable for machine learning and NLP tasks.
- Foundation for Further NLP Analysis:
  - Used in text classification, clustering, sentiment analysis, and topic modelling by providing structured data for algorithms.

Case -based Questions

1. **Imagine you are developing an application to diagnose depression in people based on their social media posts. Which application of NLP can you use to achieve this? Justify.**

Ans: For diagnosing depression based on social media posts, Sentiment Analysis (also known as Opinion Mining) is the key NLP application used. Sentiment Analysis helps in determining the emotional tone of text by analysing words, phrases, and context. It can classify posts as positive, negative, or neutral, and advanced models can detect emotions like sadness, hopelessness, or anxiety—which are indicators of depression. By leveraging machine learning and deep learning, the system can track patterns over time and provide insights into a person's mental health. This application is valuable for early detection, allowing timely intervention and support.

2. **Think of a situation where you have been asked to create an application that summarises news on climate change from various blogs by your company. Can NLP help you build this application? If yes, which feature of NLP will enable you to accomplish this task? Explain.**

Ans: Yes, NLP can help build an application that summarizes climate change news from various blogs. The key NLP feature used for this task is Text Summarization.

Text Summarization helps in automatically generating concise summaries while retaining essential information. It works in two ways: Extractive Summarization, which selects key sentences directly from the text, and Abstractive Summarization, which generates a new summary using natural language understanding. This feature enhances readability, saves time, and ensures users receive key insights without reading long articles.

3. **Consider that you are building a chatbot to answer FAQs (frequently asked questions) on a messaging app for a company that provides mobile connectivity services. Which type of chatbot will you use? What are the advantages that this chatbot will provide?**
   Ans: For answering FAQs about mobile connectivity services, a Rule-Based Chatbot (Retrieval-Based Chatbot) is ideal. It provides predefined responses based on keyword detection or intent matching, ensuring quick and accurate replies. This chatbot offers several advantages, including instant responses, consistent information, 24/7 availability, cost-effectiveness, and scalability. By automating customer support, it improves user experience while reducing the workload on human agents.

4. **You have been assigned a project where you have to categorise e-books according to their genre and type, like fiction, non-fiction, autobiographies, etc. Name the feature of NLP that will help you with your task. How does it work? [CBSE]**
   Ans: The NLP feature that helps in categorizing e-books by genre and type is Text Classification. It works by analysing the content of books and assigning them to relevant categories such as fiction, non-fiction, and autobiographies. The process begins with data preprocessing, where the text is tokenized, stopwords are removed, and words are stemmed or lemmatized. Next, feature extraction methods like Bag of Words (BoW), TF-IDF, or Word Embeddings convert the text into numerical form. A machine learning model (such as Naïve Bayes or SVM) or a deep learning model (like LSTMs or Transformers) is then trained using labelled e-book data. Finally, when a new book is processed, the trained model predicts its genre based on textual patterns. This automation helps in efficiently organising large e-book collections.

Unit 7: Advanced Python

A. Short answer type questions.

1. **Write a short note on Anaconda distribution.**
   Ans: Anaconda distribution is a powerful and widely used open source distribution of Python language for scientific computations, machine learning and data science tasks. It is an essential tool for data scientists, researchers and developers as it includes essential pre-installed libraries. It simplifies the process of managing software packages and dependencies.

2. **How to execute commands in Jupyter notebook?**
   Ans: Once you have launched Jupyter Notebook within your virtual environment, you can execute commands by creating and running Python code cells within a notebook.
   - Create a New Notebook or Open an Existing One.
   - Once you have a notebook open, you'll see an empty code cell where you can enter Python code. Click on the cell to select it, and then type or paste your Python code into the cell.

- After entering your Python code in a cell, you can execute it by either pressing "Shift + Enter" or clicking the "Run" button in the toolbar. This will run the code in the selected cell and display the output directly below the cell.

3. **Define membership operators.**
   Ans: Membership operators “in” and “not in” are used to check if a value exists in a list or sequence or not.

4. **What is an infinite loop?**
   Ans: In some cases, a condition in a ‘for’ or ‘while’ loop does not ever become false, hence the statements between the loop keep repeating indefinitely. This is called an infinite loop.

5. **Write the difference between the ‘else’ and ‘elif’ statements.**
   Ans. The else statement works with single conditions whereas the elif statement is used to test multiple conditions.

6. **What are nested loops?**
   Ans: A loop inside a loop is called a nested loop.

7. **Write the syntax for a nested if-else.**
   Ans:

```
if <condition1>:
   if <condition2>:
      statement(s)
   else:
      statement(s)
else:
   statement(s)
```

B. Long answer type questions.

1. **Write different ways to install Python libraries.**
   Ans: The different ways are:
   - Importing the entire library: This allows to access all functionalities of the library:
     import numpy
   - Importing the library with an alias: This imports the library with an alias name:
     import numpy as np
   - Importing all functions from a library: This imports all functions and objects from the library:
     from numpy import *

2. **Discuss the uses and applications of any four Python libraries.**
   Ans:
   - **NumPy:** Stands for Numerical Python. It is a fundamental package for numerical computations.

- **Pandas:** it is a powerful data analysis and manipulation library.
- **MatPlotLib:** Comprehensive library for creating static, animated and interactive visualisations in Python.
- **NLTK:** Natural Language Toolkit used for interaction with humans in natural language.

3. **What are the different components of Jupyter Notebook?**
   Ans: The components are:
   Menu: Located at the top of the page. It includes options to create new notebooks, open existing notebooks and save your work.
   Code cells: Used to type and execute programs.
   Markdown cells: Used for adding comments, headings and formatted text.
   Output area: It displays the output of the code. This could be in the form of text, plots, graphs etc.

4. **What do you mean by virtual environment?**
   Ans: A virtual environment in Python is like a separate workspace where you can install packages without affecting other projects or the system. It helps keep projects organized and avoids conflicts between different versions of Python libraries. For example, if one project needs NumPy 1.21 and another needs NumPy 1.19, a virtual environment lets you use both without problems.

5. **What is the role of indentation in Python?**
   Ans: In Python, indentation is used to define blocks of code instead of curly braces {} like in other languages. It helps in structuring the code and is mandatory for loops, functions, and conditionals. Incorrect indentation causes errors.

6. **What is the difference between a division operator and a floor division operator?**
   Ans:
   A division operator (/) performs division and returns a floating-point result, even if the operands are integers. If both operands are integers, the result will be a floating-point number, including the fractional part.
   Example:
   result = 7 / 2
   print(result) # Output: 3.5
   The floor division operator (//) performs division and returns the quotient of the division, rounded down to the nearest integer. It returns only whole numbers.
   Example:
   result = 7 // 2 print(result) # Output: 3

7. **What is the difference between division and modulus operators?**

   Ans: The **division (/) operator** returns the **quotient** of a division as a floating-point number.
   The **modulus (%) operator** returns the **remainder** of a division.
   Example:

```
print(10 / 3)  # Output: 3.3333 (Quotient)
print(10 % 3)  # Output: 1 (Remainder)
```

8. **Describe the use of input() and print() functions.**

   Ans:
   The input() function is used to take input from the user. It always returns the input as a string, irrespective of the type of input. To receive the input as a number, it has to be typecast. Example:

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

   The print() function prints a message or value. It converts a value into string before displaying it.
   Example:

```
print("Hello, ", name) #name is a variable in which a string has been accepted
```

9. **What is the difference between 'for' loop and 'while' loop? Explain with the help of a flowchart.**

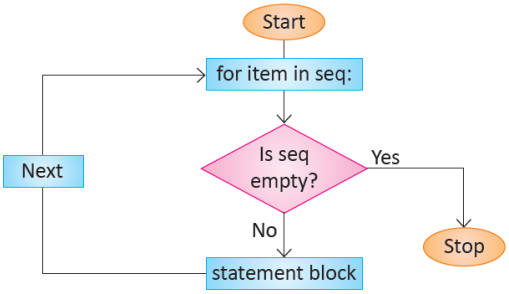
   Ans: Difference Between 'for' Loop and 'while' Loop

Feature	for Loop	while Loop
Usage	Used when the number of iterations is known.	Used when the number of iterations is unknown and based on a condition.
Structure	Iterates over a sequence (like a list, range, or string).	Runs as long as a condition is **True**.

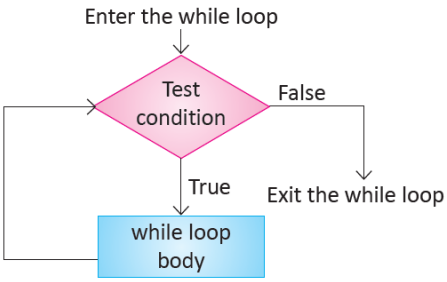
```
for i in range(5):
    print(i)  # Prints 0 to 4
```

```
i = 0
while i < 5:
    print(i)
    i += 1  # Prints 0 to 4
```

Flowchart: 'for' loop



'while' loop



10. What are Python nested if statements? Explain with an example.

Ans: Nested if statements in Python refer to if statements that are placed inside other if statements. The inner if statement gets executed only if the outer is true. An example to check if a number is non-zero and odd or even:

```
num = int(input("Enter a number: "))
# Check if the number is positive
if num >= 0:
      # Check if the number is even
      if num % 2 == 0:
      print("The number is even.")
      else:
      print("The number is odd."
```

11. What is typecasting? Explain its types.

Ans: The process of converting one data type to another is called type casting. The types are:

Implicit typecasting: The typecasting where one data type is automatically converted to another is called implicit typecasting. This makes programming simpler.

Example:

```
num1=10
num 2= 20.5
sum = num1+num2
```

Here sum is automatically converted to a float, which is the higher data type.

Explicit typecasting: In this type, the data type conversion is done manually by using int(), float() and str() functions.

Example:

```
age = int(input("Enter your age: "))
```

Here the input by the user is explicitly converted into an integer value and assigned to the variable age.

12. Write a program to find numbers divisible by 7 and multiples of 5 between 1200 and 2200.

Ans:

```
start = 1200
end = 2200
# Iterate through the range and display numbers meeting the criteria
print("Numbers divisible by 7 and multiples of 5 between 1200 and 2200:")
for num in range(start, end + 1):
    if num % 7 == 0 and num % 5 == 0:
        print(num)
```

13. Write a program to input the monthly income of an employee between 40 and 60 years old and calculate the annual income tax.

Ans:

```
# Program to calculate annual income tax for employees aged between 40
and 60

# Input employee age
age = int(input("Enter the age of the employee: "))

# Check if age is within valid range
if 40 <= age <= 60:
    # Input monthly income
    monthly_income = float(input("Enter the monthly income of the
employee: "))

    # Calculate annual income
```

```
        annual_income = monthly_income * 12

        # Determine tax based on annual income
        if annual_income <= 300000:
            tax = 0
        elif annual_income <= 500000:
            tax = annual_income * 0.05
        elif annual_income <= 1000000:
            tax = annual_income * 0.20
        else:
            tax = annual_income * 0.30

        # Output the result
        print("Annual Income: ₹", round(annual_income, 2))
        print("Annual Income Tax: ₹", round(tax, 2))

    else:
        print("This program is only for employees aged between 40 and 60.")
```

C. Predict the output of the following code snippets.

1. 200
2. *Number from* 0 *to* 99 *on a separate line.*
3. 1 2 3 4 5 6
4. FALSE
   TRUE
5. 11.0
6. 36
7. 2
8. [1, 2, 3, 5, 7]
9. 3