

General Instructions:

1. Please read the instructions carefully.
2. This Question Paper consists of **21 questions** in two sections: Section A & Section B.
3. Section A has Objective type questions whereas Section B contains Subjective type questions.
4. **Out of the given (5 + 16 =) 21 questions, a candidate has to answer (5 + 10 =) 15 questions in the allotted (maximum) time of 2 hours.**
5. All questions of a particular section must be attempted in the correct order.
6. **SECTION A - OBJECTIVE TYPE QUESTIONS (24 MARKS):**
 - i. This section has 05 questions.
 - ii. Marks allotted are mentioned against each question/part.
 - iii. There is no negative marking.
 - iv. Do as per the instructions given.
7. **SECTION B – SUBJECTIVE TYPE QUESTIONS (26 MARKS):**
 - i. This section has 16 questions.
 - ii. A candidate has to do 10 questions.
 - iii. Do as per the instructions given.
 - iv. Marks allotted are mentioned against each question/part.

Q1.	Answer any 4 out of the given 6 questions on Employability Skills (1 x 4 = 4 marks)	
i.	<p>To ensure the "Care and Maintenance" of a computer, which of the following practices should be avoided?</p> <p>A) Regularly scanning for viruses. B) Removing SPAM files. C) Blocking the cooling fans to prevent noise. D) Cleaning computer components.</p> <p>Answer: C) Blocking the cooling fans to prevent noise.</p>	1
ii.	<p>Rohit is writing an email to his boss. He ensures that he uses respectful language, correct grammar, and maintains a polite tone throughout the message. Which of the '7 Cs' of effective communication is Rohit primarily applying?</p> <p>A) Conciseness B) Concrete</p>	1

	C) Courtesy D) Clarity Answer: C) Courtesy	
iii.	Which of the following is a "Problem related to sustainable development"? A) Over-exploitation of resources and environmental degradation. B) Increased use of renewable energy. C) Conservation of indigenous knowledge. D) Use of solar cookers. Answer: A) Over-exploitation of resources and environmental degradation.	1
iv.	"The ability to work independently" is best characterized by a student who: A) Waits for the teacher to provide every instruction before starting. B) Ignores deadlines and submits work when they feel like it. C) Prefers working only in large groups to share the workload. D) Executes tasks in a specific period with no help or directives Answer: D) Executes tasks in a specific period with no help or directives.	
v.	How does entrepreneurship primarily benefit society? A) It reduces the need for technology. B) It creates jobs and contributes to economic development. C) It discourages self-employment. D) It focuses solely on personal hobbies. Answer: B) It creates jobs and contributes to economic development.	
vi.	In the sentence "The quick brown fox jumps over the lazy dog," the word "jumps" functions as which part of speech? A) Noun B) Adjective C) Verb D) Article Answer: C) Verb	1
Q2.	Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)	
i.	In the "4Ws Problem Canvas" used during Problem Scoping, which block helps you define the context or the situation in which the problem arises? A) "Who" Block B) "What" Block C) "Where" Block D) "Why" Block Answer: C) "Where" Block	1

ii.	<p>Kapil has collected a dataset of 1,000 images of cats and dogs. He notices that 950 images are of dogs and only 50 are of cats. If he proceeds without fixing this, what issue will he face during Data Exploration?</p> <p>A) Data Imbalance B) Data Scarcity C) Data Privacy violation D) Model Overfitting</p> <p>Answer: A) Data Imbalance</p>	1
iii.	<p>Deepfakes are AI-generated videos that make people appear to say or do things they never did. Which major ethical concern do Deepfakes primarily violate?</p> <p>A) Data Imbalance B) Veracity and Misinformation C) Environmental Sustainability D) Job Displacement</p> <p>Answer: B) Veracity and Misinformation</p>	1
iv.	<p>Which of the following is NOT a method of Data Exploration?</p> <p>A) Creating graphs and charts to visualize trends. B) Calculating the mean and median of the dataset. C) Training the neural network algorithms. D) Identifying outliers in the data.</p> <p>Answer: C) Training the neural network algorithms.</p>	1
v.	<p>"The Digital Divide" in the context of AI Ethics refers to:</p> <p>A) The gap between Rule-based and Learning-based AI. B) The gap between those who have access to technology/internet and those who do not. C) The difference in speed between two supercomputers. D) The division of data into training and testing sets.</p> <p>Answer: B) The gap between those who have access to technology/internet and those who do not.</p>	1
vi.	<p>An AI system is programmed with a fixed set of strict instructions: <i>"If it is raining, turn on the wiper. If it is sunny, keep wipers off."</i> This is an example of:</p> <p>A) Machine Learning Approach B) Deep Learning Approach C) Rule-Based Approach D) Unsupervised Learning</p> <p>Answer: C) Rule-Based Approach</p>	1
Q3.	Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)	

i.	<p>In a Learning-Based approach, what is the primary requirement for the machine to generate a model?</p> <p>A) A flow chart of rules. B) Data. C) A mechanical robot. D) An internet connection.</p> <p>Answer: B) Data.</p>	1
ii.	<p>Which of the following scenarios is a classic example of a "Classification" problem?</p> <p>A) Identifying whether an email is "Spam" or "Not Spam". B) Predicting the temperature for tomorrow (e.g., 34.5°C). C) Grouping customers based on purchasing behavior. D) Teaching a robot to walk by trial and error.</p> <p>Answer: A) Identifying whether an email is 'Spam' or 'Not Spam'.</p>	1
iii.	<p>In a Neural Network, where does the actual processing and feature extraction happen before the final decision is made?</p> <p>A) Input Layer B) Hidden Layers C) Output Layer D) Data Layer</p> <p>Answer: B) Hidden Layers</p>	1
iv.	<p>In the context of the AI Project Cycle, at which stage does the calculation of metrics like Accuracy and F1 Score take place?</p> <p>A) Data Exploration B) Modelling C) Evaluation D) Deployment</p> <p>Answer: C) Evaluation</p>	1
v.	<p>A fire alarm system rings when there is NO fire. What type of error is this in the language of the Confusion Matrix?</p> <p>A) True Positive B) False Negative C) False Positive D) True Negative</p> <p>Answer: C) False Positive</p>	1
vi.	<p>"Transparency" in model evaluation refers to:</p> <p>A) Making the computer screen see-through. B) The ability to explain why a model made a specific decision or prediction. C) Sharing the user's passwords with everyone. D) Using open-source software only.</p> <p>Answer: B) The ability to explain why a model made a specific decision or prediction.</p>	1

Q4.	Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)	
i.	The goal when evaluating an AI model is to: A) Maximize error and minimize accuracy B) Minimize error and maximize accuracy C) Focus solely on the number of data points used D) Prioritize the complexity of the model Answer: B) Minimize error and maximize accuracy	1
ii.	You are building an AI to detect a deadly virus. Which error is the most dangerous and must be minimized at all costs? A) False Positive (Healthy person marked as infected) B) True Negative (Healthy person marked as healthy) C) True Positive (Infected person marked as infected) D) False Negative (Infected person marked as healthy) Answer: D) False Negative	1
iii.	A developer trains an AI model using a dataset of 1,000 images. To evaluate its performance correctly, they decide to use the "Train-Test Split" method. What is the primary reason for testing the model on a separate "Testing Set" rather than the data it was trained on? A) To ensure the model can generalize its learning to new, unseen data. B) To increase the speed of the training process. C) To check if the model can memorize the training images perfectly. D) To make the dataset larger. Answer: A) To ensure generalization to unseen data.	1
iv.	In the healthcare sector, Computer Vision is primarily used for: A) Managing patient billing records. B) Translating medical prescriptions C) Scheduling doctor appointments. D). Analyzing X-rays and MRI scans to detect diseases like tumors. Answer: D) Analyzing X-rays and MRI scans to detect diseases like tumors.	1
v.	What is the smallest unit of a digital image? A) Kernel B) Filter C) Pixel D) Layer Answer: C) Pixel	1

vi.	<p>If you zoom into a digital image significantly, it becomes blurry and blocky. This phenomenon occurs because:</p> <p>A) The computer screen is broken.</p> <p>B) The image is composed of a finite number of pixels.</p> <p>C) The colors are fading.</p> <p>D) The AI is failing to process it.</p> <p>Answer: B) The image is composed of finite pixels.</p>	1
-----	---	---

Q5.	Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)	
i.	<p>Object Detection differs from Image Classification because Object Detection:</p> <p>A) Only tells you what is in the image.</p> <p>B) Tells you <i>what</i> is in the image and <i>where</i> it is.</p> <p>C) Colors every pixel of the object.</p> <p>D) Removes the background.</p> <p>Answer: B) Tells you what is in the image and where it is.</p>	1
ii.	<p>A "Smart Bot" differs from a "Script Bot" because a Smart Bot:</p> <p>A) Uses Artificial Intelligence and Machine Learning to learn from interactions.</p> <p>B) Is cheaper to build.</p> <p>C) Can only answer "Yes" or "No".</p> <p>D) Requires a human to type the answers manually.</p> <p>Answer: A) Uses AI and ML to learn from interactions..</p>	1
iii.	<p>Words like "is", "the", "and", "a" are often removed during text processing because they add little unique meaning. These words are called:</p> <p>A) Keywords B) Stem Words</p> <p>C) Rare Words D) Stop Words</p> <p>Answer: D) Stop Words</p>	1
iv.	<p>Which NLP application is used to automatically scan a long document and pull out the most important terms or names?</p> <p>A) Keyword Extraction B) Text Normalization</p> <p>C) Sentiment Analysis D) Script Bot</p> <p>Answer: A) Keyword Extraction</p>	1
v.	<p>What distinguishes lemmatization from stemming?</p> <p>A) Lemmatization produces meaningful words after affix removal, while stemming does not.</p>	1

	<p>B) Lemmatization is faster than stemming.</p> <p>C) Stemming ensures the accuracy of the final word.</p> <p>D) Stemming generates shorter words compared to lemmatization.</p> <p>Answer: A) Lemmatization produces meaningful words after affix removal, while stemming does not.</p>	
vi.	<p>Assertion: Converting the entire text into lowercase following stop word removal is a crucial preprocessing step in natural language processing.</p> <p>Reasoning: This process ensures uniformity in word representation, preventing the machine from treating words with different cases as distinct entities, thereby enhancing the accuracy of subsequent text analysis.</p> <p>A) Assertion is true, but Reasoning is false.</p> <p>B) Assertion is false, but Reasoning is true.</p> <p>C) Both are true, and Reasoning is the correct explanation of the Assertion.</p> <p>D) Both are true, but Reasoning is not the correct explanation of the Assertion.</p> <p>Answer: D) Both are true, but Reasoning is not the correct explanation of the Assertion.</p>	1

SECTION B: SUBJECTIVE TYPE QUESTIONS

Answer any 3 out of the given 5 questions on Employability Skills (2 x 3 = 6 marks)

Answer each question in 20 – 30 words.

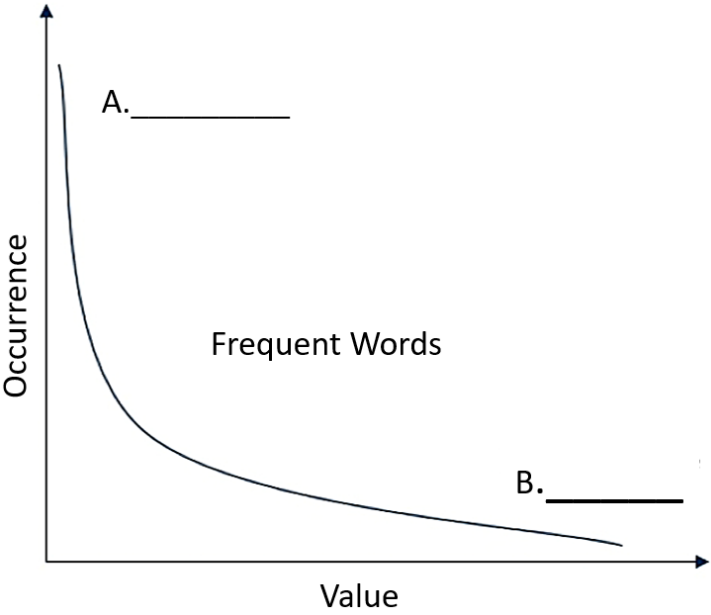
Q6.	<p>List your favorite stress management technique and elaborate why you find it the most effective.</p> <p>Technique: Physical Exercise (e.g., Yoga, Walking, or Sports).</p> <p>I find physical exercise most effective because it helps release endorphins (natural painkillers and mood elevators) in the brain. It shifts my focus away from stressful thoughts, improves sleep quality, and leaves me feeling physically relaxed and mentally rejuvenated.</p> <p>Or Any other stress management technique</p>	2
Q7.	<p>How is a computer file system similar to our physical file system in a school?</p> <p>Answer: A computer file system is analogous to a school's physical filing system in the following ways:</p> <ul style="list-style-type: none"> Storage: Just as physical files are stored in cabinets, digital data is stored in files on a hard drive. 	2

	<ul style="list-style-type: none"> • Organization: Physical files are organized into folders and labelled shelves. Similarly, digital files are organized into directories (folders) and sub-directories for easy retrieval and management. 	
Q8.	<p>Write down the common communication barriers you may come across when you move to a new city or country.</p> <p>Answer: Common barriers include:</p> <ul style="list-style-type: none"> • Linguistic Barriers: Difficulty in understanding the local language, dialect, or accent. • Cultural Barriers: Differences in social norms, gestures, and etiquette which may lead to misunderstandings. • Interpersonal Barriers: Lack of familiarity and trust with new people, leading to hesitation in communication. 	2
Q9.	<p>List the ways in which an entrepreneur affects a society.</p> <p>Answer: An entrepreneur affects society by:</p> <ul style="list-style-type: none"> • Creating Jobs: Generating employment opportunities for the community. • Economic Growth: Contributing to the national income and wealth creation. • Solving Problems: Creating innovative products or services that address societal needs and improve the standard of living. 	2
Q10.	<p>Explain the importance of education towards sustainable development.</p> <p>Answer: Education is critical for sustainable development because:</p> <ul style="list-style-type: none"> • Awareness: It spreads awareness about environmental challenges like climate change and resource depletion. • Skill Development: It equips individuals with the 'Green Skills' and knowledge needed to develop eco-friendly technologies and practices. • Responsible Citizenship: It fosters values that encourage people to act responsibly towards the planet and future generations. 	2

Answer any 4 out of the given 6 questions in 20 – 30 words each (2 x 4 = 8 marks)

Q11.	<p>"AI is a tool, not a replacement for human judgment." Discuss this statement with reference to the ethical concern of 'Human Rights and Accountability'.</p> <p>Answer: AI systems lack human qualities like empathy, moral reasoning, and distinct accountability.</p> <ul style="list-style-type: none"> • Human Rights: Relying solely on AI for high-stakes decisions (e.g., criminal sentencing or hiring) can lead to violations of human rights due to algorithmic bias. 	2
-------------	--	----------

	<ul style="list-style-type: none">Accountability: An AI cannot be held legally responsible for its actions. Therefore, a 'Human-in-the-loop' is essential to review decisions, ensuring accountability and fairness.											
Q12.	<p>Why are Convolutional Neural Networks (CNNs) preferred for Computer Vision tasks?</p> <p>Answer: CNNs are preferred for Computer Vision because:</p> <ul style="list-style-type: none">Spatial Features: Unlike regular neural networks, CNNs can automatically detect and preserve spatial hierarchies (patterns) in images, such as edges, shapes, and textures.Efficiency: They use 'convolutional layers' (filters) that scan the image, making them highly efficient at processing grid-like data (pixels) without losing the spatial relationship between them.	2										
Q13.	<p>Match the Evaluation Metric in Column A with its correct Definition / Significance in Column B.</p> <table><tr><th>Column A (Metric)</th><th>Column B (Definition / Significance)</th></tr><tr><td>1. Accuracy</td><td>A). The ratio of correctly predicted positive observations to the total predicted positive observations (Reflects the "Quality" of predictions).</td></tr><tr><td>2. Precision</td><td>B). The harmonic mean of Precision and Recall, used to balance the trade-off between them.</td></tr><tr><td>3. Recall</td><td>C). The ratio of total correct predictions (both positive and negative) to the total number of cases in the dataset.</td></tr><tr><td>4. F1 Score</td><td>D). The ratio of correctly predicted positive observations to all observations in the actual positive class (Reflects the "Quantity" found).</td></tr></table> <p>Answer: Correct Match-</p> <p>1. Accuracy — C</p> <p>2. Precision — A</p> <p>3. Recall — D</p> <p>4. F1 Score — B</p>	Column A (Metric)	Column B (Definition / Significance)	1. Accuracy	A). The ratio of correctly predicted positive observations to the total predicted positive observations (Reflects the "Quality" of predictions).	2. Precision	B). The harmonic mean of Precision and Recall, used to balance the trade-off between them.	3. Recall	C). The ratio of total correct predictions (both positive and negative) to the total number of cases in the dataset.	4. F1 Score	D). The ratio of correctly predicted positive observations to all observations in the actual positive class (Reflects the "Quantity" found).	2
Column A (Metric)	Column B (Definition / Significance)											
1. Accuracy	A). The ratio of correctly predicted positive observations to the total predicted positive observations (Reflects the "Quality" of predictions).											
2. Precision	B). The harmonic mean of Precision and Recall, used to balance the trade-off between them.											
3. Recall	C). The ratio of total correct predictions (both positive and negative) to the total number of cases in the dataset.											
4. F1 Score	D). The ratio of correctly predicted positive observations to all observations in the actual positive class (Reflects the "Quantity" found).											
Q14.	<p>Differentiate between "Classification" and "Regression" models in Supervised Learning.</p> <p>Answer:</p> <p>Classification: Used when the output variable is a category or discrete label. (Example: Predicting if an email is 'Spam' or 'Not Spam').</p>	2										

	Regression: Used when the output variable is a continuous numerical value. (Example: Predicting the price of a house or tomorrow's temperature).	
Q15.	<p>What is meant by 'Feature Extraction' in the context of Computer Vision?</p> <p>Answer: Feature Extraction is the process of identifying and isolating specific, meaningful attributes or patterns within an image, such as edges, corners, blobs, or textures. This step reduces the complexity of the image data, allowing the AI model to focus on the most relevant information for analysis.</p>	2
Q16.	<p>Consider the graph, it is a plot of the occurrence of words versus their value in a corpus in context of NLP. Suggest appropriate labels A, B for respective section of the graph.</p>  <p>Answer:</p> <p>Label A: High-frequency words or stop words</p> <p>Label B: Low-frequency or rare words</p>	2

Answer any 3 out of the given 5 questions in 50– 80 words each (4 x 3 = 12 marks)

Q17.	<p>A school wants to install an AI-powered "Smart Gate" that automatically opens only for students wearing the correct school uniform and ID card. Use the 4Ws framework to scope this problem.</p> <p>Answer:</p> <ul style="list-style-type: none"> • Who: The stakeholders are the Students (users), School Administration (beneficiaries), and Security Guards. • What: The problem is the manual effort required to check every student's uniform and ID, leading to queues or unauthorized entry. The nature of the problem is Security and Efficiency. • Where: The context/location is the School Entrance Gate. • Why: To automate security, ensure discipline regarding uniforms, prevent unauthorized access, reduce the workload on security staff, and automate the entry process. 	4
-------------	--	----------

Q18.	<p>You have a dataset of 1,000 different flowers mixed together without any name tags (Labels). You want an AI to organize them into groups based on how they look (size, color, petal shape).</p> <p>A) Which type of Machine Learning algorithm would you use?</p> <p>B) Why is this not a Supervised Learning task?</p> <p>C) Name the specific sub-category of the algorithm.</p> <p>Answers:</p> <p>A) I would use an Unsupervised Learning algorithm.</p> <p>B) This is not Supervised Learning because the dataset is unlabeled (there are no name tags/answers provided). The AI is not trained on 'correct answers' but must find patterns on its own.</p> <p>C) The specific sub-category is Clustering.</p>	4
Q19.	<p>Identify the type of model (classification, regression, clustering), are the following cases studies most likely based on?</p> <p>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.</p> <p>B) A real estate agency wants to predict the selling price of houses based on various features such as size, location, number of bedrooms, and bathrooms. They have a dataset containing historical sales data.</p> <p>C) A marketing company wants to segment its customer base into distinct groups based on purchasing behavior for targeted marketing campaigns. They have a dataset containing information such as purchase history, frequency of purchases, and amount spent.</p> <p>D) A hospital wants to group patients based on similarities in symptoms and test results to better understand disease subtypes. They have a dataset containing medical records, lab values, and symptom profiles.</p> <p>Answers:</p> <p>A. Bank (Loan Default/Non-Default): Classification (Predicting distinct categories).</p> <p>B. Real Estate (House Price): Regression (Predicting a continuous numerical value).</p> <p>C. Marketing (Customer Segmentation): Clustering (Grouping data based on behavior without predefined labels).</p> <p>D. Hospital (Grouping Patients/Disease Subtypes): Clustering (Finding hidden groups or patterns in data).</p>	4
Q20.	<p>Anshu has designed a model to detect any fraudulent transactions with credit card. The outcomes are given in the table for 10 transactions. For the given data:</p>	4

Transaction ID	Actual Value	Predicted Value
1	Non-Fraudulent	Non-Fraudulent
2	Non-Fraudulent	Fraudulent
3	Non-Fraudulent	Non-Fraudulent
4	Fraudulent	Non-Fraudulent
5	Fraudulent	Fraudulent
6	Non-Fraudulent	Non-Fraudulent
7	Fraudulent	Non-Fraudulent
8	Non-Fraudulent	Fraudulent
9	Non-Fraudulent	Non-Fraudulent
10	Non-Fraudulent	Non-Fraudulent

A) Draw the confusion matrix based on the above information.

B) Calculate the accuracy of this classification model. Show your working.

C) Write the total number of wrong predictions made by the model.

Answers:

A) Confusion Matrix:

	Predicted Fraud (T)	Predicted Non-Fraud (F)
Actual Fraud (T)	1 (TP)	2 (FN)
Actual Non-Fraud (F)	2 (FP)	5 (TN)

B) Accuracy Calculation:

Formula: $(TP + TN) / \text{Total Predictions}$

Calculation: $(1 + 5) / 10 = 6 / 10 = 0.6$

Accuracy = 60%

C) Total Wrong Predictions:

Wrong Predictions = False Positives (FP) + False Negatives (FN)

Calculation: $2 + 2 = 4$

Total Wrong Predictions = 4

Q21. Consider the following three documents:

Document 1: Akash and Ajay are best friends.

Document 2: Ajay likes football.

4

Document 3: Akash also plays football.

Perform basic text pre-processing (removing punctuation, converting to lowercase, and tokenizing) and show output after the process. Also, calculate Inverse Document frequency Vector of the corpus.

Answer:

Step 1: Text Pre-processing (Remove punctuation, Lowercase, Tokenize)

Document 1: ['akash', 'and', 'ajay', 'are', 'best', 'friends']

Document 2: ['ajay', 'likes', 'football']

Document 3: ['akash', 'also', 'plays', 'football']

Step 2: Calculate Inverse Document Frequency (IDF)

Total Number of Documents (N) = 3

Term (t)	akash	and	ajay	are	best	friends	likes	football	also	plays
Document Frequency (df)	2	1	2	1	1	1	1	2	1	1
IDF (N/df)	3/2	3/1	3/2	3/1	3/1	3/1	3/1	3/2	3/1	3/1

---O---