Question Bank

24MCAT201 - Data Science and Machine Learning

Course Type	Course Nature	CA Conduct	System	L	T	P	Credits	CA Total	CA Pass	SEE Total	SEE Pass	Total Pass
Theory	1	End Semester	Mark	2	1	0	3	50	0	50	20	50

Question Bank Summary

Sect. Part A	Sect. Part B	Easy	Med.	Chall.	Th.	Appli.
74	53	33	68	26	89	38

Part A

#	Unit	Question	COS	Categorized
1	1.1	A company hires a data analyst to extract insights from customer purchase data. Define Data Science and explain its role in solving such business problems.	CO1	Easy - Understanding - T
2	1.1	A bank receives raw customer data with incomplete records and inconsistencies. What is data preparation, and why is it critical before applying predictive models?	CO1	Easy - Remembering - T
3	1.1	An e-commerce company finds unusually high purchase amounts in its transaction data. What are outliers, and how should they be treated during data preprocessing?	CO1	Easy - Understanding - T
4	1.1	A research dataset includes “student age” and “favorite subject.” Differentiate between continuous and categorical data types using suitable examples	CO1	Easy - Understanding - T
5	1.1	A business team is asked to conduct initial exploration on sales data. What are the key objectives of data exploration in such a scenario?	CO1	Medium - Applying - A
6	1.1	A programmer is building a fraud detection system. Compare traditional programming with machine learning approaches for this problem, highlighting strengths and limitations of each	CO1	Medium - Analysing - T
7	1.1	A dataset for student performance prediction has missing exam scores. What is a missing value in data preprocessing, and how can it be handled effectively?	CO1	Easy - Understanding - T
8	1.1	A data scientist wants to summarize customer data before visualization. Explain how data exploration is divided into descriptive statistics and data visualization, and illustrate with an example.	CO1	Challenging - Applying - A
9	1.1	A healthcare startup wants to implement AI-driven diagnostics but is unsure whether to adopt AI systems or machine learning models. Compare and contrast AI and Machine Learning in terms of scope, capability, and application in healthcare	CO1	Medium - Analysing - T
10	1.2	A business team is working with a large sales dataset and wants to communicate insights to stakeholders who are not familiar with raw data. What is data visualization, and how does it help transform complex datasets into meaningful insights for decision-making?	CO1	Easy - Understanding - T
11	1.2	A marketing analyst wants to present the percentage contribution of different product lines (electronics, clothing, and groceries) to total sales. What kind of data is best represented with a pie chart, and why is this chart suitable for visualizing proportions?	CO1	Medium - Understanding - T

12	1.2	A sales manager wants to analyze company revenue and identify which product categories (electronics, furniture, clothing, etc.) contribute the most. Which visualization tool or chart would you recommend to represent this information effectively, and how would it help in decision-making?	CO1	Medium - Applying - A
13	1.2	An education researcher is investigating whether there is a relationship between students' study hours and their exam performance. Which visualization method would best highlight the correlation between the two variables, and why is this method more effective than other visualization options?	CO1	Medium - Applying - A
14	1.2	A project manager is preparing a performance report for a company's board members. They need to decide whether to use tables, graphs, or dashboards. What is the purpose of data visualization, and why is it considered a critical tool in data analysis and communication?	CO1	Medium - Understanding - T
15	2.1	State and explain one difference between Training Error and Testing Error in machine learning. Illustrate why testing error is a better measure of generalization	CO2	Easy - Understanding - T
16	2.1	Provide any two real-world examples of Supervised Learning applications. Briefly explain how labeled input-output pairs are used in each case	CO2	Medium - Remembering - T
17	2.1	Define a Learning Algorithm in Machine Learning and explain its role in optimizing model parameters. Provide one example of a commonly used algorithm	CO2	Challenging - Understanding - T
18	2.1	You are building a spam email classifier. During the initial phase, you feed the algorithm with a labeled dataset of emails marked as "Spam" or "Not Spam." The algorithm adjusts its internal parameters to identify patterns in the data. What is this phase of the machine learning process called, and why is it essential before testing the model on unseen data?	CO2	Medium - Remembering - T
19	2.2	Explain why the k-Nearest Neighbors algorithm is referred to as a lazy learner. Highlight how it delays generalization until a query is made	CO2	Challenging - Understanding - T
20	2.2	A marketing dataset contains a categorical feature "Customer Type" with values like Regular, Premium, and VIP. Explain what dummy coding is and how it converts this categorical feature into numerical columns that can be used in a machine learning model	CO2	Medium - Understanding - T
21	2.2	A company wants to scale customer ratings that range from 1 to 100 for a k-NN recommendation system. Explain how Min-Max Normalization rescales the data to a fixed range (e.g., 0 to 1) and why this is useful for distance-based algorithms	CO2	Medium - Applying - T
22	2.2	Analyze the strengths and weaknesses of the k-Nearest Neighbor (k-NN) algorithm and evaluate its suitability for classification problems with different dataset characteristics.	CO2	Easy - Remembering - T
23	2.2	Define Predictive and Descriptive Models and explain how they differ in purpose. Use examples such as sales forecasting (predictive) and customer segmentation (descriptive).	CO2	Easy - Understanding - T
24	2.2	A company wants to normalize customer spending data for a machine learning model. Explain how Z-score Standardization transforms the data	CO2	Easy - Remembering - T
25	2.3	You are working as a data analyst in a healthcare company. Your task is to classify whether a patient has a particular disease based on their symptoms. The dataset is small and contains independent features. Would you recommend Naive Bayes for this problem? Explain the strengths that make it suitable and the weaknesses that may limit its performance in other contexts	CO2	Easy - Remembering - T
26	2.3	Explain the zero probability problem in Naive Bayes classification and analyze how it impacts the calculation of posterior probabilities.	CO2	Challenging - Analysing - T
27	2.3	Define a probabilistic classifier and explain how it uses probability theory to predict the most likely class for a given input.	CO2	Easy - Remembering - T
28	3.1	A school wants to predict student performance (pass/fail) using attendance and internal marks. Explain the benefits of using a decision tree for solving such problems.	CO3	Medium - Applying - A

29	3.1	A retail company wants to classify customers into high, medium, and low spenders using purchase data. They are considering different decision tree algorithms to identify the best attribute splits.Which decision tree algorithms are available to classify the given problem.	CO3	Easy - Understanding - T
30	3.1	A retail company wants to segment customers into categories like high, medium, and low spenders based on purchase history. The data needs to be split optimally at each step to make accurate classifications. Explain what a splitting criterion and a split point are in a decision tree.	CO3	Challenging - Applying - A
31	3.1	A school wants to use a decision tree to predict student performance. List and explain the strengths and weaknesses of the decision tree algorithm.	CO3	Medium - Analysing - A
32	3.1	A hospital wants to build a decision tree to classify patients for further testing. To improve model accuracy and avoid overfitting, pruning techniques are considered. Explain pre-pruning and post-pruning in decision trees and their importance	CO3	Medium - Analysing - A
33	3.1	Define entropy and information gain and explain how they help in selecting the best attribute for splitting.	CO3	Medium - Understanding - T
34	3.2	Suppose you are designing a decision tree model for predicting whether a customer gets loan or not. Explain how classification rules are generated from the decision tree and describe rule parts.	CO3	Easy - Remembering - T
35	3.2	Explain the purpose of the Separate and Conquer strategy and the 1R algorithm, and how they could be applied to generate classification rules	CO3	Challenging - Applying - A
36	3.2	Explain the fundamental roles of both the Separate and Conquer strategy and the 1R (One Rule) algorithm within the domain of classification and rule-based machine learning.	CO3	Easy - Remembering - T
37	3.3	A researcher is studying the relationship between study hours and exam scores. Explain the concept of correlation and types of correlations using this scenario.	CO3	Easy - Understanding - T
38	3.3	Describe a detailed, critical analysis of Multiple Linear Regression (MLR) as a statistical modeling technique. Specifically, delineate its primary advantages and its disadvantages	CO3	Medium - Understanding - T
39	3.3	In the context of predictive modeling and machine learning, what are the fundamental applications of regression analysis, and how does it differ from classification?	CO3	Medium - Understanding - T
40	3.3	For a dataset of household income and expenditure, applyOrdinary Least Squares (OLS) estimation to fit a regression line. Explain the interpretation of the coefficients.	CO3	Challenging - Applying - A
41	3.3	A nutritionist wants to study the effect of calorie intake on body weight. Explain the purpose of regression analysis in this context	CO3	Medium - Remembering - T
42	4.1	Provide a formal definition of an Artificial Neural Network (ANN), outlining its structure as a computational model inspired by the biological nervous system and its primary purpose in machine learning.	CO4	Easy - Understanding - T
43	4.1	Distinguish between the broader field of Machine Learning and the specific class of algorithms known as Neural Networks (or Deep Learning). Explain where ANNs fit within the ML landscape	CO4	Medium - Understanding - T
44	4.1	Enumerate and describe three distinct application domains where Artificial Neural Networks (ANNs) have demonstrated significant superiority over traditional algorithms	CO4	Medium - Understanding - T
45	4.1	A team of engineers is designing a simple neural network to recognize handwritten digits. Explain what an artificial neuron (processing unit) is in this context, and describe its functional components and how each contributes to processing input and generating output.	CO4	Medium - Applying - A
46	4.1	A data scientist is designing a model to classify emails as spam or not spam. Explain what a Perceptron is and how it works in this scenario. Illustrate your explanation with a labeled diagram showing its inputs, weights, summation function, activation function, and output.	CO4	Medium - Evaluating - T
47	4.1	Analyze the functional characteristics of the Unit Step (or Heaviside) Activation Function. What are its primary advantages and, crucially, its fundamental limitations that restrict its use in multi-layer, gradient-based learning networks?	CO4	Challenging - Analysing - T

48	4.1	Provide the mathematical definition and a graphical representation of the Rectified Linear Unit (ReLU) activation function, explicitly stating its range and domain	CO4	Medium - Understanding - T
49	4.1	Discuss the major advantages of using the ReLU activation function and detail its primary limitations	CO4	Medium - Remembering - T
50	4.1	Define the class of functions known as "Squashing Functions" detailing the mathematical property they share. Explain their necessary role in the output layer of ANNs for tasks like binary classification or probability estimation.	CO4	Challenging - Remembering - T
51	4.1	A robotics engineer is designing a neural network to control a robotic arm for assembly tasks. Explain what a Feed Forward Network is and how it processes input signals to produce output actions in this scenario.	CO4	Challenging - Applying - A
52	4.1	A researcher is developing a neural network to predict stock prices, where past outputs influence future predictions. Explain what a Feedback (Recurrent) Network is and how feedback connections affect the network's behavior.	CO4	Medium - Applying - A
53	4.1	A machine learning practitioner is training a neural network to recognize handwritten digits. Define the cost function in this context and explain how it helps measure the network's prediction errors	CO4	Easy - Understanding - A
54	4.1	Beyond training ANNs, identify and describe specific application areas or computational tasks where the Backpropagation algorithm, is essential	CO4	Medium - Understanding - T
55	4.1	Conduct a critical evaluation of the combined framework of Artificial Neural Networks trained via the Back Propagation algorithm, detailing both their primary strengths and their major weaknesses	CO4	Challenging - Evaluating - A
56	4.1	Demonstrate the standard mathematical approach used to calculate the contribution of a single output neuron to the overall Total Squared Error of the network, outlining the calculation required before Backpropagation begins	CO4	Challenging - Applying - A
57	4.2	Explain the concept of Support Vector Machine (SVM) and analyze how it can be applied to real-world classification problems such as spam filtering or medical diagnosis.	CO4	Medium - Applying - A
58	4.2	Explain the use of Support Vector Machines (SVMs) in solving classification problems and justify their application with suitable real-world examples	CO4	Medium - Evaluating - A
59	4.2	Using the concept of SVM, define a hyperplane and explain its role in classifying emails into Spam and Non-Spam categories. Draw a diagram to illustrate the hyperplane and the separation of two classes	CO4	Medium - Analysing - T
60	4.2	What is the purpose of introducing slack variables in Support Vector Machines, and how do they help in managing non-linearly separable data while balancing the trade-off between margin maximization and classification errors?	CO4	Medium - Understanding - T
61	4.2	Explain strengths and weaknesses of support vector machines (SVM) when using non-linear kernel functions, highlighting their advantages in handling complex data and potential limitations in practical applications	CO4	Medium - Understanding - T
62	4.2	Explain the concept of a Maximum Margin Hyperplane in Support Vector Machines (SVM) and its significance in achieving optimal class separation	CO4	Medium - Understanding - T
63	5.1	A marketing team is unsure about how many customer groups to create for targeted promotions. How can you guide them in choosing the number of clusters (K) in the K-Means algorithm?	CO5	Challenging - Applying - A
64	5.1	A retail company wants to segment its customers based on their purchase patterns to design personalized marketing strategies. Explain the advantages and disadvantages of using K-Means clustering in this scenario.	CO5	Easy - Remembering - T
65	5.1	A wildlife research team has collected GPS data of animal movements. How canclustering be used to discover patterns in migration routes and group similar movement paths?	CO5	Easy - Understanding - T

66	5.1	A company has a dataset containing both labeled and unlabeled records. How would you explain the **difference between clustering and classification** to help them choose the right technique?	CO5	Medium - Understanding - T
67	5.2	A hospital developed a diagnostic system for detecting a disease. How would you explain the confusion matrix to evaluate whether the system is performing well?	CO5	Challenging - Evaluating - T
68	5.2	In a spam email detection system, how would you define and explain precision and recall to measure the accuracy of the predictions?	CO5	Easy - Understanding - T
69	5.2	Detail how the F-Measure (F1-Score) is calculated and why it is considered a more robust measure for evaluating a model's performance by harmonically balancing Precision and Recall.	CO5	Medium - Evaluating - A
70	5.2	An e-commerce company is evaluating its recommendation system. How would you use ROC curves to check how well the model distinguishes between relevant and irrelevant recommendations?	CO5	Medium - Evaluating - A
71	5.2	Elaborate on the statement: 'Cross-validation is used to check how well a model generalizes to unseen data.' Specifically, describe the procedure of k-fold cross-validation, and explain how this technique helps in detecting and mitigating the issues of overfitting and selection bias in model training.	CO5	Challenging - Understanding - T
72	5.2	A medical researcher has very limited patient data. How can bootstrap sampling be used to generate reliable performance estimates of a prediction model?	CO5	Medium - Remembering - T
73	5.2	A stock market prediction model is unstable and gives different results with different samples of data. How can bagging (bootstrap aggregating) improve its performance, and what are its advantages?	CO5	Easy - Understanding - T
74	5.2	In a fraud detection project, the initial model is not accurate enough. How does the boosting process improve model performance by focusing on difficult-to-classify cases?	CO5	Medium - Understanding - T

Part B

#	Unit	Question	COS	Categorized
1	1.1	A company explores employee data before predictive modeling. What are the different stages of data exploration, and how do they support further analysis?	CO1	Medium - Applying - A
2	1.1	A university is designing a curriculum for data-driven decision-making in business, healthcare, and social sciences. Identify and analyze the different types of Data Science and explain their relevance in real-world applications	CO1	Medium - Analysing - T
3	1.1	A retail company wants to implement Data Science to improve customer experience but is unsure of the steps involved. Illustrate the Data Science process with a diagram and explain how each stage supports decision-making	CO1	Medium - Applying - T
4	1.1	A data analyst is summarizing sales data for a quarterly business review. Explain Descriptive Statistics with examples, and analyze how it supports effective data exploration and visualization	CO1	Medium - Analysing - T
5	1.1	Explain the CRISP-DM framework with a diagram, and describe how each stage contributes to solving real-world business problems.	CO1	Easy - Remembering - T
6	1.1	Explain the role of feature selection and sampling in reducing model complexity, and discuss how these steps enhance performance and efficiency	CO1	Easy - Remembering - T
7	1.1	A financial analyst is studying the relationship between customer income, spending, and debt levels. Explain the concept of central data points and correlation in multivariate exploration with examples	CO1	Challenging - Understanding - T

8	1.1	An e-commerce firm plans to build a machine learning model to predict customer churn. Explain the process of modeling with a suitable diagram, highlighting key steps such as feature selection, training, and evaluation.	CO1	Challenging - Understanding - T
9	1.2	A researcher wants to show the marks distribution of students in a class and also compare average scores across different subjects. Differentiate between a bar chart and a histogram in terms of the type of data represented, usage, and insights provided, using this example.	CO1	Medium - Understanding - T
10	1.2	A business analyst is tasked with preparing a report to present company performance metrics to senior management. The data involves trends, proportions, and category comparisons. Identify common data visualization techniques and analyze how each can be applied effectively in real-world decision-making scenarios	CO1	Medium - Remembering - T
11	1.2	A healthcare researcher is studying a dataset that includes patient age, blood pressure, cholesterol levels, and diagnosis results. Explain multivariate visualization methods and analyze how they help in exploring relationships and patterns among multiple variables in this healthcare dataset.	CO1	Medium - Understanding - T
12	1.2	A data scientist is analyzing the distribution of exam scores in a university course. The dataset consists of a single numeric variable (marks out of 100). Explain univariate visualization methods and illustrate how they help in understanding data distribution, central tendency, and outliers in such a dataset.	CO1	Medium - Understanding - T
13	2.1	List and explain differences between Machine Learning and Traditional Programming, focusing on how rules are generated. Support your answer with a suitable example.	CO2	Medium - Understanding - T
14	2.1	Define Overfitting in Machine Learning and explain why it occurs when a model memorizes noise in training data. Briefly mention one technique to overcome it.	CO2	Medium - Understanding - T
15	2.1	A company is building a machine learning model to predict customer churn and needs to manage data, identify patterns, and assess model performance. Explain the roles of data storage, abstraction, generalization, and evaluation in helping the model learn effectively, with examples from this context.	CO2	Medium - Analysing - T
16	2.2	You are building a k-NN classifier for handwritten digit recognition. If k=1 is chosen, explain one limitation of this choice and why the model may be sensitive to noise in the training data.	CO2	Medium - Analysing - T
17	2.2	A data scientist is preparing a dataset with features that have very different ranges for a k-NN model. Explain the different scaling methods that can be applied to standardize the data	CO2	Medium - Applying - A
18	2.2	Define a lazy learner in machine learning and compare it with an eager learner by providing suitable algorithmic examples.	CO2	Medium - Understanding - T
19	2.3	A company wants to classify customer emails as spam or not spam based on word occurrences. Explain how the Naive Bayes algorithm works in this scenario, highlighting its assumption of feature independence and how it calculates probabilities to make predictions	CO2	Medium - Applying - A
20	2.3	You are developing a text classifier to categorize news articles. While testing, some new words appear that were not present in the training dataset, causing zero probability for certain classes. How can applying the Laplace estimator solve this issue, and what effect does it have on the model's predictions?	CO2	Medium - Applying - A
21	2.3	Apply the concept of discretization by grouping continuous features into bins in a Naive Bayes model, and analyze how the choice of bin size impacts classification accuracy	CO2	Challenging - Applying - A
22	3.1	Provide a comprehensive, step-by-step exposition of the process used to construct a Decision Tree classifier, explicitly detailing how this procedure embodies the Divide and Conquer algorithmic paradigm. Present a simple, tabular dataset and walk through the initial two levels of the tree construction to exemplify the iterative splitting process.	CO3	Challenging - Applying - A

23	3.1	Use Decision tree algorithm to predict whether a movie would fall into one of following three categories. Use the scatter plot to represent the pattern a. Critical Success b. Mainstream Hit c. Box Office Bust	CO3	Medium - Applying - A
24	3.1	Using the core formula for Entropy, demonstrate the step-by-step mechanism for quantifying the impurity or uncertainty of a given set of data. Specifically, calculate the Entropy for a dataset comprising 10 instances, where the binary classification split is 6 examples belonging to the "Play" class and 4 examples belonging to the "Don't play" class.	CO3	Medium - Applying - A
25	3.1	Assume an initial dataset of 10 instances (7 “eat” and 3 “Don't eat”). Now, hypothesize a single feature split, A. Given the original set's Entropy, detail the formal steps required to calculate the Information Gain provided by a hypothetical split A.	CO3	Challenging - Applying - A
26	3.1	Describe the concept of Decision Tree Pruning, differentiating between Pre-Pruning and Post-Pruning strategies with example	CO3	Easy - Remembering - T
27	3.1	A data analyst observes that in a decision tree, certain attributes appear multiple times at different levels. With a suitable example, explain how *repetition* and *replication* occur in decision trees.	CO3	Medium - Understanding - T
28	3.2	Detail the Separate-and-Conquer strategy for inducing unordered classification rule sets. Describe the iterative, greedy cycle of Rule Generation (the 'Conquer' phase) and Instance Removal (the 'Separate' phase). Illustrate this process specifically through an example where the target is the positive class, "Buys Product = Yes."	CO3	Challenging - Understanding - T
29	3.2	Explain the operational mechanism of the 1R (One Rule) algorithm. Discuss its primary function in machine learning as a performance baseline and a simple model for establishing maximum interpretability. Why is its simplicity a virtue when evaluating more complex classifiers?	CO3	Medium - Understanding - T
30	3.3	Distinguish between positive, negative, and independent correlations using illustrative real-world examples. What exactly does the Pearson Correlation Coefficient (r) measure, and what two major components (calculation and interpretive principle) constitute the standard "Pearson Correlation Problem"?	CO3	Medium - Understanding - T
31	3.3	Analyze a dataset of Year of experiences and Salary and identify the dependent and independent variables in simple linear regression. Explain how each independent variable influences the dependent variable	CO3	Medium - Analysing - A
32	3.3	Using data on O-ring failures and temperature, apply simple linear regression to predict rocket failure. Plot the results graphically	CO3	Challenging - Applying - A
33	3.3	Explain the main uses of regression analysis as a strategic tool in business and decision-making. Please define its role in any three key areas	CO3	Easy - Remembering - T
34	4.1	A team of engineers is designing an Artificial Neural Network (ANN) to mimic human visual recognition. Illustrate how a biological neural network maps to an ANN, highlighting the correspondence between neurons, synapses, and signal processing	CO4	Medium - Analysing - A
35	4.1	While building a neural network for predicting customer behavior, a data scientist needs to understand its structure. Identify and explain the key components of an Artificial Neural Network in this scenario	CO4	Easy - Remembering - T
36	4.1	A developer is designing a neural network for image classification. Explain the different types of activation functions that can be used in the hidden and output layers, and discuss how each affects the network’s learning and output	CO4	Medium - Understanding - T
37	4.1	Formulate a single-layer Perceptron problem where the objective is to classify a student's outcome (Pass/Fail) based on a single numerical input (Marks). Detail the operation of this simple neuron by explicitly defining the Unit Step Activation Function	CO4	Challenging - Creating - A
38	4.1	Explain the necessity of using the Sigmoid Activation Function when the objective shifts from a binary hard limit to generating a probabilistic output (i.e., the probability of passing). Provide the mathematical definition of the Sigmoid function and explain how its bounded range (0 to 1) makes it suitable for this Pass/Fail classification scenario.	CO4	Medium - Understanding - A

39	4.1	Define the concept of Network Topology (or architecture) in the context of Artificial Neural Networks. Distinguish and illustrate with clear, labeled diagrams the structural differences between the two primary classes of ANN topology: Feed-Forward Networks and Feedback/Recurrent Networks.	CO4	Challenging - Analysing - A
40	4.1	A data scientist is designing an Artificial Neural Network for speech recognition. Explain the possible directions of information traversal in the network and how they affect signal processing.	CO4	Easy - Understanding - T
41	4.1	While training a neural network to predict house prices, the engineer must evaluate errors effectively. Identify and explain the different types of cost functions that can be used in this scenario.	CO4	Easy - Remembering - T
42	4.1	A machine learning developer is optimizing a neural network for credit card fraud detection. Explain how Backpropagation helps improve the accuracy of the network's predictions	CO4	Easy - Understanding - T
43	4.1	During one epoch of training a neural network for image classification, the model undergoes forward and backward passes. Explain the forward phase and backward phase, and describe their roles in updating the network's weights	CO4	Easy - Understanding - T
44	4.2	Given a dataset where the classes are not linearly separable, explain the method to determine the maximum margin separating the classes and discuss how kernel functions can help achieve this	CO4	Challenging - Understanding - T
45	4.2	Explain how the Kernel Trick in Support Vector Machines (SVM) can be used to classify non-linearly separable email data. Provide an example of a kernel function and describe how it transforms the data for effective classification	CO4	Medium - Understanding - T
46	4.2	What are the most commonly used kernel functions in Support Vector Machines (SVM), and how do they help in transforming non-linearly separable data into a space where it can be effectively separated? Explain their characteristics and typical applications	CO4	Easy - Remembering - T
47	4.2	Provide a detailed explanation of Multiclass Support Vector Machines (SVM), including the strategies used for extending binary SVM to handle multiple classes and their practical applications.	CO4	Medium - Understanding - T
48	5.1	A retail business wants to segment its customers into groups based on purchase patterns. How would you explain the K-Means algorithm to achieve this?	CO5	Medium - Understanding - A
49	5.1	In a project to group students by their study habits, how does the K-Means algorithm use distance to assign and update clusters during the learning process?	CO5	Medium - Understanding - T
50	5.1	A Telecom company has collected call data records. How can distance metrics (like Euclidean distance) influence the assignment of clusters in K-Means for customer segmentation?	CO5	Challenging - Understanding - T
51	5.2	Suppose you applied clustering to group documents in a library. How would you evaluate the performance of the clusters formed by your model?	CO5	Medium - Evaluating - A
52	5.2	A credit rating agency uses machine learning to decide if a customer will default on a loan (yes/no). How would you explain the concept of a binary classifier and describe five algorithms commonly used for such tasks?	CO5	Medium - Understanding - T
53	5.2	A university is developing an AI system to predict student dropout risk. What strategies can be used to improve the ML model's performance in such cases?	CO5	Easy - Remembering - T