VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

Reg. No

Final Assessment Test(FAT) - Apr/May 2025

Programme	**M.C.A.**	Semester	**Winter Semester 2024-25**
Course Code	**PMCA507L**	Faculty Name	**Prof. Prakash B**
Course Title	**Machine Learning**	Slot	**D1+TD1**
		Class Nbr	**CH2024250501728**
Time	**3 hours**	Max. Marks	**100**

Instructions To Candidates

- Write only your registration number in the designated box on the question paper. Writing anything elsewhere on the question paper will be considered a violation.

Course Outcomes

CO1: Recognize the characteristics of machine learning that makes it useful to solve real-world problems
CO2: Provide solution for classification, regression and clustering approaches in real- world applications
CO3: Gain knowledge to combine machine learning models to achieve better results
CO4: Realize methods to reduce the dimension of the dataset used in machine learning algorithms

Section - I
Answer all Questions (7 × 10 Marks)

01. A hospital conducts a medical test for few patients to diagnose whether the patient has heart disease or not. The test results are shown as below. Apply Naive Bayes classifier and determine if the new patient (Patient ID: 11) is having likelihood of heart disease or not.

Patient ID	Blood sugar level	Blood Pressure	Age	Gender	Heart Disease
1	High	Low	Adult	Male	Positive
2	Low	Low	Adult	Female	Negative
3	Low	Normal	Adult	Male	Positive
4	High	High	Senior Citizen	Male	Positive
5	Normal	High	Senior Citizen	Male	Negative
6	High	High	Adult	Female	Positive
7	Normal	Low	Adult	Male	Negative
8	Normal	Normal	Adult	Female	Negative
9	High	Low	Adult	Male	Negative
10	Low	Normal	Adult	Male	Positive
11	Normal	Low	Senior Citizen	Female	???

[10] (CO1/K3)

02. For the data provided in Question-1, apply the ID3 algorithm and illustrate the step-by-step procedure to create a classification tree and derive the inferences. Refrain the iteration till 'max_depth=2' if the result is not achieved.

[10] (CO1/K4)

03. Consider the following data which belongs to 2 classes.

Feature 1(x1)	Feature 2 (x2)	Class
2	1	Positive
3	1	Positive
5	-1	Positive
4	0	Positive
1	0	Negative
0	-1	Negative
1	1	Negative
0	0	Negative

(a) Using SVM, illustrate the step-by-step procedure to find the hyperplane equation and plot the graph for the above given data in the table. (7 marks)
(b) Identify whether the point (x1,x2) =(4,2) belongs to Positive or Negative class? (3 marks)

[10] (CO2/K3)

04. Consider the following customers' transaction data from a leading 'e-commerce' company (For eg., Amazon). Based on the given data, the objective is to find the relationships between the products, customers choice of buying products together. Perform association rule mining and provide the inference on the derived result. The support level is considered as 33% and confidence level is 60%.

1	Television	Home Theatre	TV Stand	DTH
2	Television	DTH	Home Theatre	OTT
3	TV Stand	DTH	OTT	
4	Television	TV Stand	DTH	
5	TV Stand	DTH	Groceries	
6	Television	TV Stand	DTH	Groceries
7	Television	Groceries		
8	Television	TV Stand	DTH	
9	TV Stand	DTH	Home Theatre	Groceries
10	Television	DTH	TV Stand	Male
11	Television	TV Stand	DTH	
12	Television	TV Stand	Groceries	OTT

[10] (CO2/K3)

05. Consider the Screen Size (inches) and the Battery life (Hrs.) of a smartphone.

Features	Example-1	Example-2	Example-3	Example-4
Screen Size	5.3	6.0	5.8	6.2
Battery Life	10	14	10	11

Reduce this two dimensional features into one dimension, by applying Principal Component Analysis (PCA) method.

[10] (CO3/K4)

06. A company is developing a self-driving car that can navigate through a city. The car is equipped with sensors and cameras that provide information about the environment. The company wants to use Reinforcement Learning to train the car to make decisions about which route to take and when to stop at traffic lights. Apply the Reinforcement Learning process to train the self-driving car. Identify the states and actions used to achieve the result and elaborate the process you follow to find the reward function. (7 Marks) Also, determine the optimal policy learned by the car, and how does it improve over time? (3 Marks)

[10] (CO1/K2)

07. (a) Imagine you're working on a binary classification problem to predict whether a bank customer will default on their loan. You have a test dataset with actual labels (whether the customer defaulted or not) and the predictions made by your model. Calculate Accuracy, precision, recall, F1-Score, mis-classification rate. (5 Marks)

Customer Number	Actual Label	Predicted Label
1	Default	Default
2	No Default	Default
3	Default	No Default
4	No Default	No Default
5	Default	Default

(b) Imagine you're predicting the number of bike rentals per day in a city based on various weather conditions. Evaluate the performance of the regression model using MAE, MSE, RMSE and R-Squared (R^2) error metrics. (5 Marks).

Day	Actual Rentals	Predicted Rentals
1	150	160
2	200	190
3	175	180
4	225	210
5	190	195

[10] (CO3/K5)

Section - II
Answer all Questions (2 × 15 Marks)

08. Assume that you are an ML Engineer at 'DigiSoft Solutions' in which your team is responsible for managing machine learning projects. A new client approached your company and requested to provide a ML based solution for an 'e-commerce' business requirements (For eg., Amazon or Flipkart).
(a) As a ML engineer, elaborate the complete life cycle of a ML project with suitable example. Discuss in detail about the step-by-step process involved in building the ML based projects. (5 marks).
(b) Explain the different types of learning in AI? Also, highlight the real life applications for each categories. (5 Marks).
(c) Elaborate the pre-processing steps that you carry out as part of project development. (5 Marks)

[15] (CO1/K1)

09. Consider the following table which contains the daily expenditures (In Thousands) on groceries (X1) and entertainment (X2) of 12 persons.

Person	Groceries (X)	Entertainment (Y)
1	4	9
2	6	8
3	7	7
4	8	6
5	9	5
6	8	4
7	9	4
8	10	6
9	5	5
10	4	8
11	5	7
12	4	6

(a) Apply the k-means method by considering the number of clusters (k=3) and the associated centroids are (6,8), (9,5) and (5,7). Illustrate and tabulate the step-by-step procedure of clustering (calculate upto 2 iterations) and plot the observations in a scatter diagram (8 Marks)
(b) Construct clusters based on DBSCAN clustering technique. Plot the final results (7 Marks)

[15] (CO2/K4)

BL-Bloom's Taxonomy Levels - (K1-Remembering,K2-Understanding,K3-Applying,K4-Analysing,K5-Evaluating,K6-Creating)