VIT®

Vellore Institute of Technology

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

Reg. No. : 23MCA1030

Final Assessment Test (FAT) - May 2024

Programme	M.C.A.	Semester	WINTER SEMESTER 2023 - 24
Course Title	MACHINE LEARNING	Course Code	PMCA507L
Faculty Name	Prof. Saleena B	Slot	B2+TB2
		Class Nbr	CH2023240501386
Time	3 Hours	Max. Marks	100

General Instructions:

- Write only Register Number in the Question Paper where space is provided (right-side at the top) & do not write any other details.

Answer all questions (10 X 10 Marks = 100 Marks)

01. You're given the task of analyzing sentiments on social media platforms to understand public opinion about a particular topic or brand. [10]

(a) How would you use machine learning techniques to classify and analyze social media posts from data collection to model deployment considering challenges such as noisy data and context-dependent language? (7 Marks)

(b) Explain how PAC learning principles could be used to determine the feasibility of learning. (3 Marks)

02. Design an AND gate using the Perceptron Training rule. Assume Weight 1 (w1) = 1.2 , Weight 2 (w2) = 0.6 , Bias (b) = 0 ,Learning rate =0.5 and Threshold =1. [10]

(a) Draw the initial perception of the AND network and illustrate the step-by-step procedure of training and updation and the number of epochs needed for convergence. (8 Marks)

(b) Draw the Final perceptron of the trained network with the updated weights. (2 marks)

03. A Medical Expert has 5 Training Samples for diagnosis tasks. Cases correspond to individuals based on three parameters (Fever, Vomiting, and Diarrhoea). They are rated on a scale of 1(No) to 5(very high). [10]

Classify the sample (Fever = 4, Vomiting =1, and Diarrhoea =2) using the KNN method where K=3.

Sample	Fever	Vomiting	Diarrhoea	Classification
c1	1	1	1	Healthy (H)
c2	3	1	1	Influenza (I)
c3	4	2	2	Food poisoning(F)
c4	2	1	1	Influenza (I)
c5	5	2	1	Food poisoning(F)

04. Consider the dataset below which has the details of the customers specified with their income, age, and number of credit cards used. The class attribute with Yes and No specifies if the customer is a defaulter or not. [10]

Customer	Age	Income	No. of credit cards	Class
Rohan	35	35k	3	No
Rohit	22	50k	2	Yes
Ashok	63	200k	1	No
Sam	59	170K	1	No
Alex	25	40k	4	Yes

(a) Apply the Gini Index and find the root attribute. (3 marks)

(b) Draw either the left sub-tree or right sub-tree from the root to the leaf node (7 Marks)

05. You are a data scientist working for a healthcare company that aims to improve patient outcomes by predicting the risk of developing a particular medical condition based on various health parameters. The company has provided you with a dataset containing information such as patients' demographic details, medical history, lifestyle factors, and diagnostic test results. Your task is to build an ensemble predictive model that accurately identifies patients at high risk of developing the condition, allowing healthcare providers to intervene early and provide proactive care. **[10]**

(a) Discuss any 3 appropriate ensemble classifiers and their contribution to accomplishing the task with neat diagrams (8 marks)

(b) Discuss the techniques used to combine predictions of ensemble classifiers for the above scenario. (2 Marks)

06. Consider the following table which contains the daily expenditures (In Thousands) on food (X1) and clothing (X2) of 5 persons. Apply the k-means method, assuming that the observations belong to two groups and the initial seed points of the groups are A and E. Illustrate and tabulate the step-by-step procedure of clustering (8 Marks) and plot the observations in a scatter diagram. (2 Marks) **[10]**

Person	X1	X2
A	3	2
B	4	1
C	2	5
D	5	2
E	6	1

07. Consider the following table which contains the daily expenditures (In Thousands) on food (X1) and clothing (X2) of 5 persons. Construct clusters and dendrograms(2 Marks) based on Agglomerative clustering using Average Linkage. Illustrate and tabulate the step-by-step procedure of clustering (8 Marks). **[10]**

Person	X1	X2
A	2	4
B	8	2
C	9	3
D	1	5
E	8	5

08. Consider the height and the weight features. **[10]**

Height	4.5	5.9	5.3	5.5
Weight	60	65	55	60

Reduce this 2-dimensional feature into one by performing the following

(a) Identify the Covariance Matrix (4 Marks)

(b) Find the Eigen values and Eigen Vectors (6 Marks)

09. In a school competition, student teams participate in a challenge where they navigate a robocar through a maze without direct visual feedback. Teams aim to guide their robocars efficiently through the maze, avoiding obstacles along the way. The winning team is determined by the fastest completion time. Inspired by this game, you aim to develop a computer game with embedded machine learning. Which machine learning paradigm is most appropriate for implementing this scenario? Describe the procedure or criteria for developing the game. **[10]**

10. Identify which classifiers among k-Nearest Neighbors, Logistic Regression, Naïve Bayes, Random Forests, and Support Vector Machine are suitable for predicting whether a customer will default on a loan based on financial and demographic information. The dataset consists of 10,000 samples with 70 features and 1 target variable. Justify the selection or exclusion of each classifier with an appropriate reason. (5 Classifiers * 2 marks each =10 Marks) **[10]**