VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)
CHENNAI

Reg. Number:	23MCA1030

Continuous Assessment Test(CAT) – I - February 2024

Programme	:	Master of Computer Applications	Semester	:	Winter 2023-24
Course Code & Course Title	:	PMCA507L-Machine Learning	Class Number	:	CH2023240501386
Faculty	:	Dr.B.Saleena	Slot	:	B2+TB2
Duration	:	1 ½ Hours	Max. Mark		50

General Instructions:

- Write only your registration number on the question paper in the box provided and do not write other information.
- Only non-programmable calculator without storage is permitted

Answer ALL questions

Q. No	Sub Sec.	Description	Marks
1.		The survey results of 6 online stores were taken. The relationship between the monthly e-commerce sales and the online advertising costs is given below. **Monthly sales (In lakhs)**: 368, 340, 665, 954, 331, 556 **Online Advertising cost (In lakhs)**: 1.7, 1.5, 2.8, 5, 1.3, 2.2 (a) Find the equation of the straight line that fits the data best. Illustrate the step-by-step procedure for forming the equation. (4) (b) Identify the association between them and predict the monthly sales for advertisement cost of 1.4 lakhs (6)	10
2.		The below table contains 11 actual and predicted samples of the model whether the patient has a tumour or not. **S.No**: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 **Actual**: Yes, Yes, No, Yes, No, Yes, No, No, Yes, No, Yes **Predicted**: Yes, No, No, Yes, No, Yes, Yes, No, Yes, No, Yes Create a matrix that will help us to evaluate the classification model and calculate the accuracy, precision, recall, and F1-measure metrics.	10

3. Consider the following training data set for predicting whether a loan borrower will default on their payments. Estimate conditional probabilities using an appropriate classifier for the loan classification problem. Illustrate the step-by-step procedure for the same. 15

Tid	Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125k	No
2	No	Married	100K	No
3	No	Single	70k	No
4	Yes	Married	120k	No
5	No	Divorced	95k	Yes
6	No	Married	60k	No
7	Yes	Divorced	85k	No
8	No	Single	85k	Yes
9	No	Married	75k	No
10	No	Single	90k	Yes

Classify the below test transaction data.

Tid	Owner	Marital Status	Annual Income	Defaulted Borrower
11	No	Single	80k	???

4. The company dataset given below has 10 instances. The company's profit is decided based on the age, competitive environment, and type of domain the people work on. Apply the ID3 algorithm and illustrate the step-by-step procedure to create a classification tree and derive the inferences that will help the company maintain a good profit. 15

Age	Competition	Type	Class:Profit
Old	Yes	Software	No
Old	No	Software	No
Old	No	Hardware	No
Mid	Yes	Software	No
Mid	Yes	Hardware	No
Mid	No	Hardware	Yes
Mid	No	Software	Yes
Young	Yes	Software	Yes
Young	No	Hardware	Yes
Young	No	Software	Yes

**********All the best **********