

Reg. No.: Q0BR31037

Name :



VIT
Vellore Institute of Technology
(Established in 1984)

Continuous Assessment Test-1 – August 2022

Programme	B.Tech. CSE (AIR)	Semester	Fall 2022-23
Course	Machine Learning and its Applications	Code	CSE3105
Faculty	Dr. Priyadarshini.J	Slot(s)	B2
		Class Nbr(s)	CH2022231000531
Time	1½ Hours	Max. Marks	50

Consider the dataset in Table 1, which illustrates the student's Final Assessment Test (FAT) result based on their Continuous Assessment Tests (CAT1 and CAT2) and the number of hours they invested to prepare for the assessments.

Table 1: Student Result Dataset

Student	CAT 1 Marks (50)	CAT 2 Marks (50)	Study Hours	Class: Result
1	30 ✓	35 ✓	4 ✓	Pass ✓
2	42 ✓	45 ✓	6 ✓	Pass
3	20 ✓	17 ✓	1	Fail
4	49	47	6	Fail
5	25 ✓	22 ✓	2	Pass
6	34 ✓	40 ✓	2	Pass ✓
7	45	48	6	Pass
8	17	10 ✓	0	Fail
9	25	20 ✓	1	Fail
10	35	38	3	Pass

Linear regression
100%

Multiple
100%

3.5 ✓
7.8
7.7
7.7
7.7
7.7
7.7
7.7
7.7
7.7

- 1 Identify and list any 3 machine learning models which is appropriate for the dataset in table 1 and justify your answer with explanation for each.
- 2 a) Identify how to split the data set in table 1 based on different conditions for Rows 1 to 6. [7 Marks]
- b) i. Predict the student's FAT result for the rows 7 to 10 using the designed decision tree from above [2 Marks].
For the predicted FAT result, perform the following:
 - ii. Represent correctness of model by constructing the confusion matrix. [2 Marks]
 - iii. How often is the built classifier correct? [1 Mark]
 - iv. When it's actually "yes", how often does it predict "yes"? [1 Mark]
 - v. When it's actually "no", how often does it predict "yes"? [1 Mark]
 - vi. When it predicts "yes", how often is it correct? [1 Mark]

st. made

9.7
7.7
7.7
7.7
7.7
7.7
7.7
7.7
7.7
7.7

- 3 Apply an algorithm which stores all the available data and classifies a new data point based on the similarity for the dataset (Rows 1 to 6) specified in Table 1. Predict a student's FAT result for the rows 7 to 10. (Assume k). 10
- 4 Compare the predicted value of the final assessment test result [Rows 9 and 10 of Table 1] of question number 2 and 3. Does the prediction using two different models yields the same result? If yes/no, discuss your view on the comparison. 5
- 5 Estimate the relationship between two quantitative variables. 10

Y(Output)	X(Input)
3	1
5	3
7	4
9	7
10	8

- a) Plot the values and find a linear function (a non-vertical straight line) that predicts the dependent variable values as a function of the independent variable. (7 marks)
- b) Calculate the mean squared prediction errors. (3 marks) ✓

Note : write all the formulas used

- 6 By using NLP, suppose, you detect spam e-mails in your inbox. Assume that the word 'alone' occurs in 80% of the spam messages in your account. Also, assume 'alone' occurs in 10% of your desired e-mails. If 30% of the received e-mails are considered as a spam, and if you receive a new e-mail which contains 'alone', what is the probability that it is a spam? 5

$$\textcircled{1} \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$\textcircled{2} \frac{\sum x^2 \sum y - \sum x \sum xy}{n \sum x^2 - (\sum x)^2}$$