



National Forensics Sciences University, Goa Campus
Mid- semester Examination, Oct-2024

Programme – MTECH AIDS

Sem – I

Date -08/10/2024

Subject Name: Fundamentals of Data Science and Machine Learning

Subject Code- CTMTAIDS SI P4

Time- 1.5 Hours

Max. Marks- 50

Instructions - 1) Answer all questions. 2) Assume suitable data.

Q.1 Solve any four

20 marks

- a. Suppose 40% of the students in a class are male, and 30% of the students are male and prefer online classes. What is the probability that a randomly chosen student prefers online classes, given that they are male? 5 marks
- b. Explain the Data Cleaning Techniques with example. 5 marks
- c. Write a Mathematical Model of Logistic Regression 5 marks
- d. Write Key techniques in Data Reduction. 5 marks
- e. Assume an asteroid is expected to hit Goa. As a data scientist what would be your role. 5 marks

Q.2 Attempt all

15 marks

- a. Consider the Confusion Matrix data Write the Formula and calculate the values for Accuracy, Precision, F1 Score. 5 marks

Confusion Matrix	Predicted Positive (1)	Predicted Negative (0)
Actual Positive (1)	50	10
Actual Negative (0)	5	35

- b. Explain 3 Different types of data with example. 5 marks
- c. Describe how the data aggregation and normalization work and why it is used in data transformation. 5 marks

Q.3 Attempt any one

8 marks

- A) List the key concepts in Data Integration and resolve the data integration for the Coding Club data given below.

Coding Club Data

StudentID	Name	Address	Date of Enrollment	Event Type	Party Pending
101	Vatsal Hasmukh	123 Oak St Vasyapur	8/15/2020	Workshop	5000
101	Vatsal Has Mukh	123 Oak Street Vasyapur	8/15/2020	Seminar	NULL
102	Dr Nikhil	456 Colva Beach	1/10/2021	Session	10000
103	Shivam Khatarnak	NULL	9/5/2019	Null	NULL

OR

- B) Based on the table below, calculate the support for the itemsets: {Bread}, {Bread, Milk}, {Diapers, Coke}. Calculate the confidence for the following rules: {Bread} \rightarrow {Milk}, {Diapers} \rightarrow {Coke}. Based on the support and confidence, derive useful association rules. 8 marks

Transaction ID	Items Purchased
1	{Bread, Milk, Eggs}
2	{Bread, Diapers, Coke, Eggs}
3	{Milk, Diapers, Coke, Cola}
4	{Bread, Milk, Diapers, Coke}
5	{Bread, Milk, Cola}

Q. 4 Attempt any one

7 marks

ID	Age	Income
1	23	35,000
2	45	85,000
3	36	50,000
4	52	120,000
5	27	42,000
6	60	200,000
7	33	55,000
8	48	70,000
9	29	40,000
10	41	90,000

- A) Apply Equal Width Binning and equal Frequency Binning to the given data. Divide the data into 3 equal bins and show the result.

OR

- B) You are tasked with designing a data science solution to analyse the placement data of students from your college's placement cell. Assume some data and explain. 7 marks

*** All the best***

Fundamentals of Data Science & ML