DEPARTMENT OF COMPUTER SCIENCE ENGINEERING (DATA SCIENCE)

UNIT TEST-I

Class:TEDS **Semester:V** **Subject:DWM**

Date: 05-Aug-2023 **Time:10:00 to 11:30am** **Max marks: 40**

Note the following instructions

1. **Attempt all questions.**
2. **Draw neat diagrams wherever necessary.**
3. **Write everything in Black ink (no pencil) only.**
4. **Assume data, if missing, with justification.**

Q.N	Questions	MARKS	CO	Bloom Level	PO
Q.1.	**Attempt any two.**				
A.	Data for salary analysis include 30, 36,47,50,52,52,56,60,63,70,70,110 Compute the first, second and third quartile for this data. Construct box plot for visualize.	[5]	CO2	L3	PO1
B.	Suppose a group of sales price records has been sorted as follows: 6,9,12,13,15,25,50,70,72,92,204,232. Partition them into three bins by equal frequency (Eqi-depth) partitioning method. Make use of bin mean method to perform data smoothing.	[5]	CO2	L3	PO1
C.	Identify the value of correlation coefficient from the following table	[5]	CO2	L3	PO1

Subject	Age	Glucose_Level
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

D.	Data for salary analysis include 30, 36,47,50,52,52,56,60,63,70,70,110 Apply Min-max , Z score , Decimal Scaling to normalize the data.	[5]	CO2	L3	PO1
Q.2.	**Attempt any two**				
A.	Apply the decision tree algorithm and show the generated rules for the following: A simple example of the stock market involving only discrete ranges has profit as categorical attribute, with values {up, down} and the training data set is as shown in the table.	[10]	CO3	L3	PO2

Age	Competition	Type	Profit
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	Up

B.	Apply Naive Bayes classifier algorithm for buys computer and classify the tuple X=<Age=Young, Income=Medium, Student=Yes, credit-rating=fair>	[10]	CO3	L3	PO2

ID	Age	Income	Student	Credit-rating	Buys-Com.
1	Young	High	No	Fair	No
2	Young	High	No	Good	No
3	Middle	High	No	Fair	Yes
4	Old	Medium	No	Fair	Yes
5	Old	Low	Yes	Fair	Yes
6	Old	Low	Yes	Good	No
7	Middle	Low	Yes	Good	Yes
8	Young	Medium	No	Fair	No
9	Young	Low	Yes	Fair	Yes
10	Old	Medium	Yes	Fair	Yes
11	Young	Medium	Yes	Good	Yes
12	Middle	Medium	No	Good	Yes
13	Middle	High	Yes	Fair	Yes
14	Old	Medium	No	Good	No

C.	Build confusion matrix and identify all evaluation measures for the confusion matrix for the following: Suppose in a data set of the cancer, there are total 10000 tuples, out of that 300 are positive and 9700 are negative and our model has predicated 90 are positive and 9560 are negative.	[10]	CO3	L3	PO2

Q.3.	**Attempt any one.**				
A.	College Wants to record the Marks for the courses completed by students using the dimensions: i) Course ii) Student iii) Time & a measure Aggregate marks. Construct a Cube and perform following OLAP operations. i) Rollup ii) Drill down iii) Slice v) Dice v) Pivot	[10]	CO1	L3	PO1
B.	Consider following dimensions for a supermarket chain: Product, Store, Time and Promotion. With respect to this business scenario, answer the following questions. Clearly state any reasonable assumptions you make. a) Construct the star schema b) Can you convert the star schema to snowflake schema? If yes, justify and build the snowflake schema. Clearly depict the fact table(s). Dimension tables their attributes and measures.	[10]	CO1	L3	PO1