1T01236 - T.E.(Information Technology Engineering)(SEM-VI)(Choice Base Credit Grading System) (R- 19) (C Scheme) / 37311 - Data Mining & Business Intelligence

QP CODE: 10042141 DATE: 11/12/2023

(3 Hours) **[Total Marks: 80]**

NOTE:

1. Question No 1 is compulsory
2. Attempt any three questions from remaining.
3. Assume suitable data if necessary and state the same.

Q.1 A) Explain types of attributes used in data exploration **(10)**

B) Explain DBSCAN algorithm with example. **(10)**

Q.2 A) Explain K means algorithm in detail. Apply K-means Algorithm to divide the given set of values {2,3,6,8,9,12,15,18,22} into 3 clusters **(10)**

B) Compare Bagging and Boosting of a classifier **(10)**

Q.3 A) Explain Multilevel and Multidimensional Association rules with suitable examples **(10)**

B) Using the given training dataset classify the following tuple using Naïve Bayes Algorithm: <Homeowner: No, Marital Status: Married, Job experience:3> **(10)**

Homeowner	Marital Status	Job experience (in years)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

Q.4 A) Define data mining. Explain KDD process with help of a suitable diagram **(10)**

B) For the table given perform Apriori algorithm and show frequent item set and strong association rules. Assume Minimum Support of 30% and Minimum confidence of 70%. **(10)**

TID	Items
01	1, 3, 4, 6
02	2, 3, 5, 7
03	1, 2, 3, 5, 8
04	2, 5, 9, 10
05	1, 4

Q.5 A) What is noisy data? How to handle it **(10)**
For the following data D={4,8,9,15,21,21,24,25,26,28,29,34}
Number of bins =3
Perform the following:
i. Partition into equal frequency bins
ii. Smoothing by bin means
iii. Smoothing by bin boundaries

B) Define data warehouse. Explain data warehouse architecture with help of a diagram **(10)**

Q.6 A) What is an outlier? List types of outliers. Describe methods used for outlier analysis. **(10)**

B) Design BI system for Fraud Detection? Explain all steps from data collection to decision making **(10)**
