U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

February / March 2023 Semester End Main Examinations

Programme: B.E. **Semester: V**
Branch: Information Science and Engineering **Duration: 3 hrs.**
Course Code: 20IS5PEDMG **Max Marks: 100**
Course: Data Mining **Date: 07.03.2023**

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

UNIT - I

1 a) How data mining is used in knowledge discovery process? **05**

b) Illustrate different types of data mining task. **05**

c) Assume any one example to compute the following measures of similarity and dissimilarity. **10**

i. Cosine measure
ii. Euclidean measure
iii. Manhattan measure

UNIT - II

2 a) What measures of node impurity is taken in decision tree induction. **05**

b) Give the algorithm for decision tree induction with neat steps. **05**

c) State Bayes theorem. How Naive Baye's classifier is used for classification? Explain with example. **10**

OR

3 a) Explicate Hunt's algorithm for inducing decision trees. Also write the design issues associated with it. **10**

b) What measures are used for selecting best split in decision tree. Give example. **05**

c) Write a note on Rule-based Classifiers. **05**

UNIT - III

4 a) For a given transaction data, generate frequent itemset and identify valid Associaton Rules with minimum support as 60% and minimum confience as 75%. **10**

Tid	Items
1	Bread, Cheese, Eggs, Juice
2	Bread, Cheese, Juice
3	Bread, Milk, Yogurt
4	Bread, Juice, Milk
5	Cheese, Juice, Milk

b) State Apriori Principle. Briefly explain, Apriori algorithm for Frequent. Itemset generation. **10**

OR

5 a) With an example, explain maximal, frequency Itemset and closed frequent Itemset techniques for compact representation. **10**

b) Define FP Tree. Construct FP Tree for given transaction data with neat steps: Data Set **10**

Tid	Items
1	{a, b}
2	{b, c,d}
3	{a, c, d, e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

UNIT - IV

6 a) Why Cluster analysis is required? Briefly explain different types of cluster analysis methods. **10**

b) What is density based clustering? How DBSCAN algorithm is used for clustering. **10**

UNIT - V

7 a) Explain chameleon, Multiphase hierarchical clustering using dynamic modeling. **10**

b) Compare the following types of clustering techniques and state which clustering is best? **10**

i) Fuzzy Clustering
ii) Density based Clustering
iii) Graph based Clustering.

****