U.S.N | | | | | | | | | | |

P.E.S. College of Engineering, Mandya - 571 401

(An Autonomous Institution affiliated to VTU, Belagavi)

First Semester, M. Tech - Computer Science and Engineering (MCSE)

Semester End Examination; Jan. -2020

Advances in Data Mining

Time: 3 hrs *Max. Marks: 100*

***Note**: Answer **FIVE** full questions, selecting **ONE** full question from each unit.*

UNIT - I

1 a. What is data mining? Discuss the major issues in data mining. 10

b. Explain the different kinds of data and patterns can be mined. 10

2 a. List and explain OLAP operations that can be performed on multi dimensional data. 10

b. What is a data warehouse? Discuss the benefits of implementation of a data warehouse. 10

UNIT - II

3 a. Illustrate the frequent itemset generation using Apriori algorithm. 10

b. Construct FP tree for the following data which has five transactions. Explain in detail.

TID	Item
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}
4	{a, d, e}
5	{a, b, c}

10

4 a. Explain with examples, how clustering and nearest neighbor are used for prediction? 10

b. Explain in brief for estimating prediction accuracy of classification methods and its improving accuracy of classification methods. 10

UNIT - III

5 a. Explain different types of clustering with examples. 10

b. Explain the K-means algorithm with example. 10

6 a. Discuss the partitioning and hierarchical clustering methods with examples. 10

b. Explain the density based outlier detection with example. 10

UNIT - IV

7 a. Explain the following:

i) Time series data

ii) Histograms

iii) Spatial data

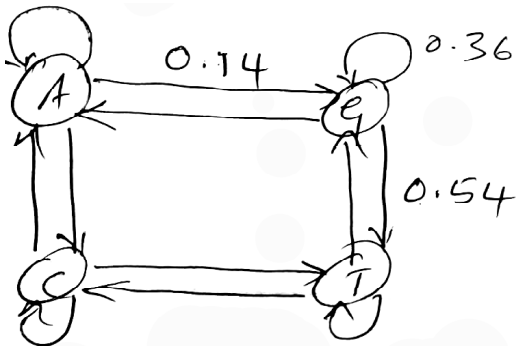
iv) Sketches

v) Random sampling

10

Contd…2

b. Explain Markov chain model and calculate how probable is X in the given below model for CPG islands where the states of the model are A, G, G and T.



10

8 a. Explain trend analysis with all its components. 10

b. Discuss how mining of sequence patterns in biological data can be achieved? 10

UNIT - V

9 a. Explain basic measures for text retrieval. Calculate the TF-IDF value of a term in a document for t_4 in documents d_5 in the given term frequency matrix.

document / term	t_1	t_2	t_3	t_4	t_5	t_6
d_1	0	5	0	8	10	4
d_2	32	0	0	16	7	19
d_3	0	17	9	4	0	0
d_4	22	3	0	5	12	15
d_5	0	9	12	0	2	4

10

b. Discuss how data mining can be achieved on temporal and spatial databases? 10

10 a. Discuss the dimensions and measures of a spatial data cube with example. 10

b. Explain the following:

i) Similarity search 10

ii) Multi dimensional analysis

* * *