



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

(An Autonomous Institution affiliated to VTU, Belgaum)

First Semester, M. Tech - Computer Engineering (MCEN)

Semester End Examination; Jan/Feb - 2016

Data Warehousing and Mining

Time: 3 hrs

Max. Marks: 100

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

UNIT - I

- 1 a. Define Data warehouse. Write any five differences between OLTP and OLAP systems. 7
- b. With suitable example explain different types of schemas used in multi dimensional model. 9
- c. Define Data cube, also discuss three measures used in multidimensional data. 4
- 2 a. What is data mining? Explain the challenges that motivated the development of Data mining. 10
- b. Explain the process of knowledge discovery in data base (KDD) with neat diagram. 6
- c. Discuss various tasks of data mining. 4

UNIT - II

- 3 a. Define classification. Explain the general approach for solving a classification problem with neat diagram. 10
- b. Explain the various measures for selecting the best splits with an example. 10
- 4 a. Discuss general-to-specific and specific-to-general rule growing strategies, give suitable example. 8
- b. Write the algorithm for K-nearest neighbor classifier. 4
- c. Discuss the methods used to measure the performance of a classifier. 8

UNIT - III

- 5 a. Consider the following transaction dataset,

Tid	1	2	3	4	5
Items	{a, b}	{b, c, d}	{a, c, d, e}	{a, d, e}	{a, b, c}

6	7	8	9	10
{a, b, c, d}	{a}	{a, b, c}	{a, b, d}	{b, c, e}

Construct the FP tree. Show the trees separately after reading each transaction.

- b. Explain the various properties of objective measures. 6
- c. Write a note on sequential pattern discovery. 4
- 6 a. Develop the Apriori algorithm for the generating frequent item set generation. 8
- b. Which are the factors affecting the computational complexity of apriori algorithm. Explain them. 6
- c. Explain the method of candidate generation and pruning with an example. 6

UNIT - IV

- 7 a. Define cluster analysis? What are the different types of clusters? 10
- b. Differentiate between ; 6
- i) Hierarchical versus partitional.
- ii) Complete versus partial.
- c. Write DBSCAN algorithm. 4
- 8 a. How can the generalization be performed on set-valued and list-valued attributes? Give example. 10
- b. Explain the different dimensions and measures used in a spatial Data cube. 10

UNIT - V

- 9 a. What is multimedia DB? Explain the methods used in similarity based retrieval in image DB based on image signature. 10
- b. Explain text retrieval indexing technique with suitable example. 6
- c. List two basic measures for assessing the quality of text retrieval. 4
- 10 a. Discuss the use of DM application for telecom Industry. 10
- b. What are the trends in DM. 10

* * *