



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

(An Autonomous Institution affiliated to VTU, Belagavi)

Fourth Semester, Master of Computer Applications (MCA)

Semester End Examination; June - 2017

Data Warehousing and Data Mining

Time: 3 hrs

Max. Marks: 100

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

UNIT - I

- 1 a. How did Inmon and Imhoff define operational Data stores? Draw a simple ODS structure and explain ODS design and Implementation. 10
- b. Define data cube. With suitable example, explain the various data cube operations. 10
- 2 a. What is data warehousing? What are the guidelines for implementing data warehousing? 10
- b. Explain star-schema architecture of multinational data model of data warehouse. How it is different from snow-flatter schema? 10

UNIT - II

- 3 a. What is data mining? With neat diagram, explain the process of Knowledge discovery in database. 10
- b. Discuss the various challenges that motivate the development of data mining. 5
- c. What is attributes? Explain the different types of attributes. 5
- 4 a. What is pre-processing? What are the steps involved in it? Explain any two steps in detail. 10
- b. Briefly explain data mining tasks. 5
- c. What are the different data sets supported by data mining? 5

UNIT - III

- 5 a. What is an association rule discovery? Write down the formal definition of support and confidence. 5
- b. Calculate the SMC, Jaccard and cosine similarity for the following binary vector and document vector: 5
 - i) $x = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ $y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$
 - ii) $x = (3, 2, 0, 5, 0, 0, 0, 2, 0, 0)$ $y = (1, 0, 0, 0, 0, 0, 0, 1, 0, 2)$
- c. Write an algorithm for frequent item set generation using Apriori method with an example. 10
- 6 a. Draw the FP-tree for the following transaction data set. Find the frequent item set using FP-growth algorithm. 10

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}

b. Explain the alternative methods for generating frequent item sets in detail. 10

UNIT - IV

7 a. What is classification? Explain Hunt’s algorithm and illustrate its working. 10

b. How to build a rule based classifier using direct method for rule extraction? Also explain sequential covering algorithm with an example. 10

8 a. Design a general approaches for solving a classification problem. Demonstrate with an example. 10

b. State and explain Bayer’s theorem. Explain with an example the Bayer’s theorem for classification. 10

UNIT - V

9 a. Define clustering. Explain different types of clustering. 10

b. What are the different approaches used to detect the anamoly in data sets? 10

10 a. What is cluster analysis? Explain the features of cluster analysis. 10

b. Discuss the following clustering algorithm with example : 10

i) K-means

ii) DBSCAN.

* * * *