**U.S.N**

# P.E.S. College of Engineering, Mandya - 571 401
*(An Autonomous Institution affiliated to VTU, Belgaum)*
**Seventh Semester, B.E. - Information Science and Engineering**
**Semester End Examination; Dec - 2016/Jan - 2017**
**Data Mining**

*Time: 3 hrs* *Max. Marks: 100*

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

## UNIT - I

| | | |
|---|---|---|
| 1 a. | Why a data warehouse is separated from operational databases? Give two reasons. | 4 |
| b. | Define metadata. Explain what does metadata repository contains? | 8 |
| c. | Draw and explain a three-tier data warehousing architecture. | 8 |
| 2 a. | List and explain types of OLAP server and functions performed by OLAP. | 10 |
| b. | Explain MOLAP composes and list out advantage and disadvantage of MOLAP. | 10 |

## UNIT - II

| | | |
|---|---|---|
| 3 a. | What is data mining? Explain the process of knowledge discovery in databases. | 10 |
| b. | Explain the different types of attributes with example. | 5 |
| c. | Mention any four real world examples that describe the activities of data mining task. | 5 |
| 4 a. | What is data pre-processing? Explain different types of sampling and dimensionality reduction techniques. | 10 |
| b. | What are the three standard approaches to feature selections? Describe the flow chart of a feature subset selection process. | 10 |

## UNIT - III

5 a. Explain frequent Item set generation using Apriori principle/algorithm. 10

b. Given the following set of transaction. Find the frequent item sets for the transaction with min sup = 0.5

| $T_{ID}$ | List of item |
|---|---|
| $T_1$ | $I_1, I_2, I_5$ |
| $T_2$ | $I_2, I_4$ |
| $T_3$ | $I_2, I_3$ |
| $T_4$ | $I_1, I_2, I_4$ |
| $T_5$ | $I_1, I_3$ |
| $T_6$ | $I_2, I_3$ |
| $T_7$ | $I_1, I_3$ |
| $T_8$ | $I_1, I_2, I_3, I_5$ |
| $T_9$ | $I_1, I_2, I_3$ |

10

6 a. How can we improve the efficiency of Apriori based mining? Explain the techniques. 10

b. Find the frequent item sets for the following database with sup min = 2

| $T_{ID}$ | Items |
|------|---------|
| 10 | A, C, D |
| 20 | B, C, E |
| 30 | A, B, C, E |
| 40 | B, E |

10

## UNIT - IV

7 a. How does classifier work? Explain how rule based classifier works with an example.  10

b. Explain characteristics of rule based classifier and k-nearest neighbour classification algorithm.  10

8 a. Consider football games between two rival teams (0/1), suppose team (0) wins 65% of time and team (1) wins the reaming matches. Among the games won by team (0) only 30% of these come from playing as team (1)'s football field. On the other hand 75% of the victories for team (1) are obtained while playing at home. If team (1) is to host the next match between two teams, which team will most likely emerge as the winners?  10

b. State and explain Bayesian Belief Network and Algorithm for generating the topology of a Bayesian Network.  10

## UNIT - V

9 a. What do you mean by clustering? Discuss any two clustering techniques with explain.  12

b. What are outliers? Explain with an example.  8

10. Write a short note on ;

a) Web mining

b) Data processing

c) DBSCAN

d) Decision trees.  20

✳ ✳ ✳