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

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; May / June - 2018
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. Define Star-schema and Snowflake schema. Design Star-schema and Snowflake schema of a data warehouse for sales containing sales as a fact table with time, item, branch and location as dimension table. Assume suitable attributes for each.   10

  b. Differentiate between OLTP and OLAP.   5

  c. Write the syntax for cube definition and dimension definition.   5

2 a. Discuss all OLAP operations in the multidimensional data along with example.   10

  b. Discuss guidelines for data warehouse implementation.   10

UNIT - II

3 a. What is Data mining? Explain the process of Knowledge Discovery in Databases (KDD) with a neat diagram.   8

  b. Define simple matching coefficient and Jaccard coefficient with the equation.   8

  c. Explain the different characteristics of datasets.   4

4 a. Explain any five data mining applications.   10

  b. Explain different types of attributes with an example and list the properties that describe attributes.   10

UNIT - III

5 a. Write and discuss the Apriori algorithm for the frequent item set generation.   10

  b. Explain alternative methods for generating frequent item set generation.   10

6 a. What is FP-tree? Construct an FP-tree for following transaction data along with steps involved in the construction :

TID	Items
1	{a, b}
2	{b, c, d}
3	{a, c, d, c}
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}

10

   b. Explain the factor that affects the computational complexity of the Apriori algorithm.   10

UNIT - IV

7 a. Explain general approach for solving classification problem along with performance metrics.   8

   b. Explain the attribute test conditions for binary, nominal, ordinal and continuous.   12

8 a. Write the algorithm for decision tree and explain the steps of algorithm.   12

   b. Explain the steps for Naive Bayesian classifier.   8

UNIT - V

9 a. Write the DBSCAN algorithm and explain time and space complexity in DBSCAN.   10

   b. Explain agglomerative and divisive hierarchical clustering with an example.   6

   c. Write the steps in K-means algorithms.   4

10 a. Explain the requirement of clustering in data mining.   10

   b. Explain different types of cluster analysis methods.   10

✳ ✳ ✳ ✳