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

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

(An Autonomous Institution affiliated to VTU, Belagavi)

Sixth Semester, B.E. - Computer Science and Engineering

Semester End Examination; May / June - 2019

Data Warehouse and 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 a Data Warehouse? Explain the four keywords that distinguish Data Warehouse from other Data Repository. 10

b. What is a Data Cube? Write the various distinguish features between OLTP and OLAP. 10

2 a. How are concept hierarchies useful in OLAP? With example, explain OLAP operations on multidimensional data. 10

b. With a neat diagram, illustrate and analyze three-tier Data Warehouse architecture. 10

UNIT - II

3 a. What is Data Mining? Explain data mining as a step in the process of knowledge discovery. 10

b. Discuss the various approaches to data cleaning as a process. 10

4 a. Explain the following dimensionality reduction techniques : 10

i) Wavelet Transform ii) Principal Components Analysis

b. Describe the various methods of data integration and transformation. 10

UNIT - III

5 a. Analyze and write an algorithm for inducing a decision tree from training tuples. 10

b. Explain the following attribute selection measures. 10

6 a. What are Bayesian classifiers? Explain the steps of Naive Bayesian classifiers. 10

b. List and explain the techniques for accessing accuracy based on randomly sampled partitions of a given data. 10

UNIT - IV

7 a. Analyze and write Apriori algorithm for discovering frequent itemsets. 10

b. Describe the various methods for improving the efficiency of Apriori algorithm. 10

8 a. A database has five transactions, let minimum support = 60% and minimum confidence = 80% 10

TID	Items-bought
T_{100}	{M, O, N, K, E, Y}
T_{200}	{D, O, N, K, E, Y}
T_{300}	{M, A, K, E}
T_{400}	{M, U, C, K, Y}
T_{500}	{C, O, O, K, I, E}

Find all frequent itemsets using FP growth.

Contd…2

b. Explain with an example, the following items :

i) Frequent itemsets

ii) Support and Confidence

iii) Association rules

10

UNIT - V

9 a. What is Cluster Analysis? Explain the requirements of Clustering in data mining. 10

b. Write the following algorithms :

i) K-means partitioning algorithm 10

ii) K-medoides partitioning algorithm

10 a. Briefly describe the two methods of Hierarchical Clustering. 10

b. With an example, explain density based method. 10

* * * *