UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

End Semester Examination, December 2017
Program: B.tech CSE with specialization CCVT,OSS, Mainframe technology, Cyber-Law
Semester – VII
Subject (Course): Data ware housing and Data Mining   **Max. Marks: 100**
Course Code   : CSEG 411   **Duration : 3 Hrs**
No. of page/s:

SECTION A (Attempt all Questions)

1. Explain following terms:
   - a) Roll up
   - b) Drill down
   - c) Slice
   - d) Dice
   - e) Probability density function with 2 examples   (5)
2. Define data mining primitives in web mining.   (5)
3. Define K-NN and Decision tree algorithm by giving examples.   (5)
4. Compare and contrast classification and clustering.   (5)

SECTION B (Attempt all Questions)

5. Explain the Apriori algorithm. Also explain how the association rules are generated from frequent item sets.   (10)
6. Explain the differences between supervised and unsupervised learning. Explain K-Means algorithm in detail. (10)
7. Discuss the term 'data cleaning'. What are different methods for handling the missing values?   (10)
8. Explain the different data cube operations   (10)

SECTION C (Attempt any two Questions)

9. What are the differences between the three main types of data ware house usages: Information processing, Analytical processing and Data mining?   (20)

10. What is decision tree method? Explain how classification is done using decision tree algorithm? (20)

11. Explain web mining with its primitive description and how it is related to parallel SQL execution? (20)

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

End Semester Examination, December 2017
Program: B.tech CSE with specialization CCVT,OSS, Mainframe technology, Cyber-Law
Semester – VII
Subject (Course): Data ware housing and Data Mining   **Max. Marks: 100**
Course Code   : CSEG 411   **Duration : 3 Hrs**
No. of page/s:

SECTION A (Attempt all Questions)

1. What is descriptive and predictive data mining?   (5)
2. What are the advantages of dimensional modelling?   (5)
3. Define K-NN and Decision tree algorithm by giving examples.   (5)
4. Explain web mining and how knowledge discovery is performed over web content?   (5)

SECTION B (Attempt all Questions)

5. Describe various stages in building a data warehouse with a neat and describe.   (10)
6. (a). Explain different types of cluster analysis techniques.   (5)
    (b). Discuss the outlier analysis.   (5)
7. Discuss the term 'data cleaning'. What are different methods for handling the missing values?   (10)
8. Explain with diagrammatic illustration as a data mining as a step in the process of knowledge discovery   (10)

SECTION C (Attempt any two Questions)

9. Write short notes on following: (20)

(a) Data Transformation
(b) Applications of Data Mining
(c) OLAP vs OLTP
(d) Data Mining Task Primitives

10. Explain the steps of the "Apriori Algorithm" for mining frequent item sets. ? Apply Apriori algorithm in following data set   (20)

Trans ID	Items Purchased
101	Apple, Orange, Litchi, Grapes
102	Apple, Mango
103	Mango, Grapes, Apple
104	Apple, Orange, Litchi, Grapes
105	Pears, Litchi
106	Pears
107	Pears, Mango
108	Apple, Orange, Strawberry, Litchi, Grapes
109	Strawberry, Grapes
110	Apple, Orange, Grapes

11. Discuss the challenges in mining the World Wide Web, and explain the role of search engines by showing neat diagram of its query execution process (20)