Seventh Semester B.E. Degree Examination, June/July 2015
Data Ware Housing and Data Mining

Time: 3 hrs.
Max. Marks:100

Note: *Answer any FIVE full questions, selecting atleast TWO questions from each part.*

PART – A

1 a. Explain the characteristics of ODS. **(06 Marks)**
b. List the major steps involved in the ETL process. **(06 Marks)**
c. Based on oracle, what are difference between OLTP and data warehouse systems. **(08 Marks)**

2 a. Discuss the FASMI characteristics of OLAP. **(05 Marks)**
b. Explain Codd's OLAP rules. **(10 Marks)**
c. Describe the difference between ROLAP and MOLAP. **(05 Marks)**

3 a. What is data preprocessing? Explain various data preprocessing tasks. **(14 Marks)**
b. Explain the following :
   i) Euclidean distance
   ii) Simple matching coefficient
   iii) Jaccard coefficient. **(06 Marks)**

4 a. Explain frequent itemset generation in the apriori algorithm. **(10 Marks)**
b. What is FP – Growth algorithm? In what way it is used to find frequency itemsets? **(03 Marks)**
c. Construct the FP tree for following data set. Show the trees separately after reading each transaction.

Tid	1	2	3	4	5
Items	{a, b}	{b, c, d}	{a, c, d, e}	{a, d, e}	{a, b, c}
Tid	6	7	8	9	10
Items	{a, b, c, d}	{a}	{a, b, c}	{a, b, d}	{b, c, e}

(07 Marks)

PART – B

5 a. What is classification? Explain the two classification models with example. **(06 Marks)**
b. Discuss the characteristics of decision tree induction algorithms. **(10 Marks)**
c. Explain sequential covering algorithm in rule –based classifier. **(04 Marks)**

6 a. List five criteria for evaluating classification methods. Discuss them briefly.   **(05 Marks)**

    b. What is predictive accuracy of classification methods? Explain different types of estimating the accuracy of a method.   **(07 Marks)**

    c. Consider the following training set for predicting the loan default problem :

Tid	Home owner	Marital status	Defaulted borrower	Annual income
1	Yes	Single	No	125 k
2	No	Married	No	100 k
3	No	Single	No	70 K
4	Yes	Married	No	120 k
5	No	Divorced	Yes	95 k
6	No	Married	No	60 k
7	Yes	Divorced	No	220 k
8	No	Single	Yes	85 k
9	No	Married	No	75 k
10	No	Single	Yes	90 k

      Find the conditional independence for given training set using Bayes theorem for classification.   **(08 Marks)**

7 a. List and explain the desired features of cluster analysis.   **(08 Marks)**

    b. Explain the K – means clustering algorithm with suitable examples.   **(12 Marks)**

8   Write short notes on :

    a. Web content mining

    b. Unstructed text

    c. Text clustering

    d. Temporal data mining tasks.   **(20 Marks)**

* * * * *