USN ☐☐☐☐☐☐☐☐☐☐

13MCA442

Fourth Semester MCA Degree Examination, June/July 2016
Data Warehousing and Data Mining

Time: 3 hrs.

Max. Marks:100

Note: *Answer any FIVE full questions.*

1 a. Define a datawarehouse. List and define the key features of a datawarehouse. **(07 Marks)**
   b. With a neat labeled diagram briefly explain a multi-tiered architecture of a data warehouse. **(07 Marks)**
   c. Explain briefly the schemas for multi–dimensional data models. **(06 Marks)**

2 a. Define data mining. With a neat diagram explain the KDD process. **(08 Marks)**
   b. List and explain the different data mining techniques used during the data preprocessing. **(12 Marks)**

3 a. List and explain the types of attributes with respective to qualitative and quantitative measurements. **(07 Marks)**
   b. List and explain the different data mining applications. **(07 Marks)**
   c. Consider :
   $$x = (1, 0, 0, 0, 1, 1, 1, 1)$$
   $$y = (0, 1, 1, 0, 1, 0, 0, 1).$$
   Evaluate simple matching coefficient (SMC) and Jaccord coefficient (JC). **(06 Marks)**

4 a. State Apriori principle for generating itemsets that are frequent. Construct itemset lattice for itemset $I = \{I_1, I_2, I_3, I_4\}$ and list all the itemsets subsets. **(10 Marks)**
   b. State the FP–growth algorithm. Construct FP–tree for the following transaction data set :

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

5 a. Define a decision tree induction. Give Hunt's algorithm. Construct a decision tree for the following data set : **(10 Marks)**

Instance	a_1	a_2	Target class
1	T	T	+
2	T	T	+
3	T	F	−
4	F	F	+
5	F	T	−
6	F	T	−

   b. Write the K–nearest neighbor classification algorithm. State and explain the characteristics of nearest neighbor classifiers. **(10 Marks)**

6   a.   List and explain the other evaluation criteria for classification methods.   (10 Marks)

     b.   With an example, explain the multiclass problem.   (10 Marks)

7   a.   Define a clustering technique. List and briefly explain the applications.   (07 Marks)

     b.   Define Bisecting K–means algorithm. Explain the same for a four clusters.   (06 Marks)

     c.   State the CLIQOE. Explain its strengths and limitations.   (07 Marks)

8   a.   List and explain outlier (anomaly) detection methods.   (07 Marks)

     b.   Give the relative density outlier score algorithm. Explain with an example.   (07 Marks)

     c.   List and explain the issues related to outlier.   (06 Marks)

* * * * *