AUTUMN END SEMESTER EXAMINATION-2017

7th Semester B.Tech

DATA MINING

IT-4003

(Regular-2014 & Back of Previous Admitted Batches)

Time: 3 Hours Full Marks: 60

Answer any SIX questions including question No.1 which is compulsory.

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable and all parts of a question should be answered at one place only.

1. Answer all the questions. [2 × 10]

   (a) What is k-nearest neighbour classification?

   (b) What is locality of reference in DDBs?

   (c) What is supervised classification? How it is different from clustering?

   (d) Give a comparative analysis of Centralized Database System and Distributed database system?

   (e) "Data mining is a multidisciplinary field". Justify your answer.

   (f) How to find interestingness of an association rule?

   (g) What is ETL Process? How it is helpful in preprocessing the data?

   (h) Describe various primitives for specifying a data mining task.

   (i) What is outlier? How to handle outlier?

   (j) What is data mart?

2. (a) What is KDD Process? Explain briefly how the knowledge is extracted from it. [4]

(b) Explain the major functions of data mining in multimedia and how it can be used for surveillance system? [4]

3. (a) Discuss about Bayesian classification and explain how Bayesian approach applied on data sets. [4]

(b) Describe various statistical measures for data mining. [4]

4. (a) Describe the k-means and k-mediods algorithm in terms of shape of cluster that can be determined and parameter that must be specified? [4]

(b) What is genetic algorithm? Explain the cross over process briefly? [4]

5. (a) What is if-then rule? What are the parameters used to find the accuracy of the rule? [4]

(b) What is Classification? Explain an example of decision tree. [4]

6. (a) What is clustering? Differentiate between Hierarchical and Partitional Clustering. [4]

(b) Define Spatial Data Mining. What is the application of Spatial Data Mining in real time? [4]

7. (a) A database has five transactions. [4]

Let min_support=50% and min_confidence=75%.

TID	Items_bought
T100	Bread,Cheese,Eggs,Juice
T200	Bread,Cheese,Juice
T300	Bread,Milk,Yoghurt
T400	Bread, Juice, Milk
T500	Cheese,Juice,Milk

Find all frequent item sets using Apriority Algorithm.

(b) What is the difference between Euclidean distance and Manhattan distance? Given two objects represented by the tuples (22,1,42,10) and (20,0,36,8). [4]

I. Compute the Euclidean distance between two objects.

II. Compute the Manhattan distance between two objects.

8. Write Short notes on *any four*. [2 × 4]

(a) Web Data mining

(b) Confusion Matrix

(c) Temporal Data mining

(d) Rough Set

(e) Neural Network classifier

– ***** –