SPRING END SEMESTER EXAMINATION-2015

2nd Semester M.Tech / Ph.D

DATAMINING AND DATAWAREHOUSING (CS-6301)

(Regular-2014 Admitted Batch)

Full Marks: 60 **Time: 3 Hours**

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 the following questions: [2 × 10

a) List out the functionality of metadata.

b) Give the difference between the Horizontal and Vertical Parallelism.

c) What are the steps to be followed to store the external source into the data warehouse?

d) What is Correlation Analysis?

e) What do you mean by segmentation in Discretization of Numerical value?

f) Define conditional pattern base.

g) Give the main advantage in employing a lazy learning method.

h) What are the two data structures in cluster analysis?

i) Name any two Grid-Based clustering methods.

j) What is sequential pattern mining?

2. a) Explain various methods of data cleaning in detail. [4

b) Diagrammatically illustrate and discuss the data warehousing architecture with briefly explain components of data warehouse. [4

3. a) How data mining system can be integrated with a data warehouse? Discuss with an example. [4

b) Write different approaches to data transformation. Explain various methods of data cleaning with examples. [4

4. BIRCH and CLARANS are two interesting clustering algorithms that perform effective clustering in large data sets. [4+4

(i) Outline how BIRCH performs clustering in large data sets.

(ii) Compare and outline the major differences of the two scalable clustering algorithms BIRCH and CLARANS.

5. a) With relevant example discuss constraint based cluster analysis. [4

b) Explain the major function of data mining in multimedia and how it can be used for surveillance system. [4

6. A database has four transactions. Let min sup=60% and min conf=80%. [8

TID	DATE	ITEMS BOUGHT
T100	10/15/07	{K,A,B}
T200	10/15/07	{D,A,C,E,B}
T300	10/19/07	{C,A,B,E}
T400	10/22/07	{B,A,D}

Find all frequent item sets using Apriori and FP growth, respectively. Compare the efficiency of the two mining process.

7. a) Discuss about Bayesian classification and explain how Bayesian approach applied on Datasets. [4

   b) Describe the various descriptive statistical measures for data mining. [4

8. Write Shortnotes: (Any Two) [4 × 2

   a) Confusion Matrix

   b) Bit Mapped Indexing

   c) Different techniques for Feature Extraction

   d) Support Vector machine

x x x x x