AUTUMN END SEMESTER EXAMINATION-2015

1st Semester M.Tech

DATA MINING & DATA WAREHOUSING (CS-6301)

(Regular-2015 & Back of Previous Admitted Batches)

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: $[2 \times 10$

 (a) Define the features of density based clustering method.

 (b) Explain the term variance and standard deviation with example.

 (c) What is Jaccard coefficient? Explain with example.

 (d) Explain correlation analysis of nominal attributes using χ^2 with example.

 (e) Explain the term ETL with respect to data warehouse.

 (f) Define support and confidence with respect to the given example.
      Example: computer => antivirus software [support = 5%, confidence = 60%].

 (g) What is tree pruning? Explain with example.

 (h) Define the term coverage and accuracy with respect to Rule-Based Classification.

 (i) Define requirement for Cluster Analysis.

 (j) List out any two commercial data mining tools.

(1)

2. (a) Describe and explain major issues in data mining. [4

   (b) A database has five transactions. Let min_support = 60% [4
   and min_confidence=80%

Transaction ID	Item Bought
T1	{Mango, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T2	{Doll, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T3	{Mango, Apple, Key-chain, Eggs}
T4	{Mango, Umbrella, Corn, Key-chain, Yo-yo}
T5	{Corn, Onion, Onion, Key-chain, Ice-cream, Eggs}

3. (a) Define Decision Tree Induction. Explain any two attribute [4
   selection method.

   (b) Explain Classification by Back propagation with suitable [4
   example.

4. (a) Define and explain constrained based cluster analysis with [4
   suitable example.

   (b) Given is the sample data set: [4

Outlook	Tempe-rature	Humidity	Windy	Class (Play Tennis)
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Overcast	Hot	High	False	P
Rain	Mild	High	False	P
Rain	Cool	Normal	False	P
Rain	Cool	Normal	True	N
Overcast	Cool	Normal	True	P
Sunny	Mild	High	False	N
Sunny	Cool	Normal	False	P
Rain	Mild	Normal	False	P
Sunny	Mild	Normal	True	P
Overcast	Mild	High	True	P
Overcast	Hot	Normal	False	P
Rain	Mild	High	True	N

(2)

Predict the class label of the new sample given below using naïve Bayesian

Classifier.

Sunny	Cool	High	False	?

5. (a) Explain the term data value conflict detection and resolution. Explain the method of dimension reduction in which irrelevant, weakly relevant or redundant attributes or dimensions are detected and removed.   [1+3

   (b) Explain K-Means algorithm with suitable example.   [4

6. (a) What is data transformation? Explain the data transformation by normalization with example.   [1+3

   (b) What is data warehouse? Explain the schemas for multidimensional data model.   [1+3

7. (a) Explain the requirement of normalization in data mining.   [4

   (b) Explain the typical OLAP data cube operations to materialize different views.   [4

8.   Write short notes (Any two)   [4 × 2

   a) Support vector machine

   b) OLAP & OLTP

   c) FP – Tree

   d) Genetic Algorithm

X X X X X