Code No: **R41053** **R10** **Set No. 1**

IV B.Tech I Semester Supplementary Examinations, February - 2019

DATA WARE HOUSING AND DATA MINING

(Common to Computer Science and Engineering and Information Technology)

Time: 3 hours **Max. Marks: 75**

Answer any FIVE Questions
All Questions carry equal marks
*****

1 a) Illustrate the role played by sourcing, acquisition, cleanup and transformation tools in building a data warehouse. [8]
b) How does data mining differ from Knowledge Discovery in Databases? [7]

2 Explain the following:
a) Cosine similarity
b) Jaccard coefficient
c) Correlation. [15]

3 Suppose that a data warehouse for *Big University* consists of the following four dimensions: student, course, semester, and instructor, and two measures count and avg grade. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg grade measure stores the actual course grade of the student. At higher conceptual levels, avg grade stores the average grade for the given combination.
a) Draw a snowflake schema diagram for the data warehouse.
b) Starting with the base cuboid [student; course; semester; instructor], what specific *OLAP* operations (e.g., roll-up from semester to year) should one perform in order to list the average grade of *CS* courses for each Big University student.
c) If each dimension has five levels (including all), such as "student < major < status < university < all", how many cuboids will this cube contain (including the base and apex cuboids)? [15]

4 a) Explain the algorithm for constructing a decision tree induction from training samples. [9]
b) What do you mean by Cross validation? Explain briefly. [6]

5 a) Explain about Bayesian Theorem. [7]
b) What is the need for classification? Discuss briefly. [8]

6 a) What is a Time-series database? State four major components or movements for characterizing time-series data. [8]
b) Explain the need of *support and confidence* for finding frequent item sets. [7]

7 a) Briefly outline how to evaluate the accuracy of a classifier or Predictor. [10]
b) Write and explain the algorithm for mining frequent item sets without candidate generation. [5]

8 a) How agglomerative hierarchical clustering works? Explain with an example. [8]
b) List the strengths and Weaknesses of DBSCAN algorithm. [7]