Exam.Code:0918
Sub. Code: 33457

2055
B.E. (Computer Science and Engineering)
Sixth Semester
Elective – I
CS-605C: Data Mining and Analysis

Time allowed: 3 Hours **Max. Marks: 50**

NOTE: Attempt five questions in all, including Question No. 1 (Section-A) which is compulsory and selecting two questions each from Section B-C.

x-x-x

Section -A

Q 1(a) Illustrate the main difference between OLAP and OLAM (10)

(b) Define Support and Confidence metrics. How would you interpret 5% Support and 80% confidence.

(c) Highlight the main limitations of Apriori algorithm.

(d) What is the size of fact table with dimensions Time (Two records per week for 5 years), Location (500 cities), Product (one record per brand for 400 brands having 5000 products).

(e) List two applications highlighting the use of Time Series data.

Section -B

Q2 (a) List the different data mining functionalities. How over-fitting of Data mining model is a major data mining issue. (5)

(b) Describe the role of ETL layer in Data Warehouse architecture. What factors determine the periodicity of ETL process? (5)

Q3 (a) What is starnet query model? How it helps in creating aggregated fact model? (5)

(b) What is data cleaning? Describe the different ways to deal with the noisy data. (5)

Q 4 (a) How would you calculate the Entropy of an attribute? How this can be used to calculate the relevance of attributes? (5)

(b) How Class Comparison and correlation analysis is performed. (5)

Section -C

Q5 (a) What is constraint based association rule mining? Describe the different constraints and their relation using different examples. (10)

(b) Illustrate the process of FP-Growth algorithm and show how only two database scans are sufficient for candidate generation.

Q6 (a) What is Gini Index? What is the complexity of decision tree algorithm? How performance of decision tree algorithm can be improved? (5)

(b) Differentiate between lazy and eager learners? What is the type of k-NN algorithm? Describe the distance weighted nearest neighbour algorithm. (5)

Q7 (a) Describe the EM Algorithm? Explain the role of both steps using an example? (5)

(b) What are Spatial databases? Describe the different clustering approaches used in spatial databases. (5)

x-x-x