Exam.Code:0918
Sub. Code: 6798

1058

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. I which is compulsory and selecting two questions from each Unit.*

x-x-x

I. Attempt the following:-

a) List the main steps in knowledge discovery process.

b) Why do we use surrogate keys in data warehouse?

c) What are fact less fact tables?

d) Why data warehouses are time variant but non-volatile.

e) What is convertible constraints?

f) Why support is not sufficient parameter to measure the success of Frequent Pattern mining?

g) What is Zero probability problem in Bayesian classifiers?

h) How difference between two ordinal variables is measured?

i) What are time series databases?

j) What is divisive clustering? (10x1)

UNIT – I

II. a) What is Online Analytical mining? Explain its architecture in detail.

b) Differentiate among the different materialization of data cubes. (6,4)

III. a) Why do we use aggregates in data warehouse? Explain the aggregate fact table using an example.

b) What are different ways to measure the dispersion of data? Explain the use of Quantile plots. (5,5)

IV. a) How redundancy is handled during data integration? How correlation analysis on numerical and categorical data is performed?

b) Describe different OLAP operations that are performed on data warehouses? (5,5)

P.T.O.

UNIT – II

V. Why FP-growth algorithm is better than Apriroi? Explain the process of generating FP-tree and conditional patterns base using an example? (10)

VI. a) What is the complexity of Decision tree classifier? How can we avoid overfitting in decision trees? How ensemble of classifiers can help?

b) What is partitioning around mediods? How is it better than k-means algorithm? (6,4)

VII. a) Illustrate the difference between web usage, content and structure mining.

b) What is dissimilarity matrix? How we differentiate between good or bad clustering? (5,5)

x-x-x