Exam.Code:0918
Sub. Code: 6569

2054

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) Why do we say, "Data warehouse is Subject Oriented and time invariant"? (10)

(b) What are limitations of using production data keys?

(c) What are virtual data warehouses?

(d) Highlight key issues in implementing Apriori Algorithm.

(e) List two applications highlighting the use of Multimedia data mining.

Section -B

Q2 (a) What are the different elements in OLAM architecture? Highlight the roles of different layers in it. (5)

(b) What are the different types of indexing used in data warehouses? Consider a city attribute in database to store 20 different cities as strings of 20 characters. If we use bitmap index for this attribute, then how much space can we save for 1 million records? (5)

Q3 (a) Following are the number of complaints registered by a service centre in last 30 days, (5)

120, 240, 350, 300, 430, 520, 530, NaN, 250, 210, 290, 210, 390, 670, 130, 30, 310, 400, 600, NaN. 120, 210, 370, 490, 220, 400, 380, 340, 210, 320

Pre-process the data and find first and third quartile of data, show the boxplot and quantile plot, discretize the data into 3 levels.

(b) Suppose that a data warehouse consists of the three dimensions *time, doctor*, and *patient*, and the two measures *count* and *charge*, where *charge* is the fee that a doctor charges a patient for a visit. (5)

(a) Enumerate three classes of schemas that are popularly used for modeling data warehouses.

(b) Draw a schema diagram for the above data warehouse.

(c) Starting with the base cuboid [*day, doctor, patient*], what specific *OLAP operations* should be performed to list the total fee collected by each doctor in 2024?

Q 4 (a) What are different ways to materialize cubes. How many cuboids will be materialized in full materialization for base cuboids having dimensions location (2 levels), time (4 levels) and product(8 levels). (5)

(b) Explain the data transformation approaches. Highlight the main issues and corresponding solutions to overcome. (5)

Section -C

Q5 What is frequent pattern mining? How FP Growth algorithm overcome the limitations of Apriori algorithm. Illustrate the construction of rules using FP growth algorithm using an example. (10)

Q6 (a) What is Overfitting? How overfitting in decision tree algorithm can be prevented? Describe some remedies and discuss their effect on outcome. (5)

(b) What is constraint based association rule mining? Explain different types of constraints. (5)

Q7 (a) Why Naïve Bayesian classifier is called so? What is the role of 'likelihood' and 'prior' in it. Explain its working using an example. (5)

(b) What are sequence databases? Explain the different mining approaches in sequence databses using an example. (5)

x-x-x