Exam.Code:0918
Sub. Code: 6413

2053

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) What is Schema Integration? (10)

(b) Differentiate between t-weights and d-weights.

(c) What are the metrics used to compare the performance of Classification Algorithm?

(d) What is data discretization?

(e) What is difference between time series and sequence database?

Section -B

Q2 (a) Illustrate the differences between ROLAP, MOLAP and HOLAP. Which architecture is preferred in large organization (5)

(b) What is partial materialization of data warehouse? Why Full materialization is not feasible in large dimensional data? Explain by taking a suitable example. (5)

Q3 (a) Following are the number of customers visited the store in last 30 days, (5)

250, 350, 415, 200; 230, 420, 500, NaN, 450, 50, 340, 260, 90, 470, 530, 60, 380, 440, 560, NaN. 420, 310, 170, 190, 290, 470, 320, 450, 310, 220

Pre-process the data and find first and third quartile of data, show the boxplot and quantile plot, divide them into equal size 3 bins and smooth these by boundary.

(b) What is the purpose of Aggregate Fact table? What are the advantages of using these? (5)

Q 4 (a) How data cubes are computed? Describe the different operations that can be performed on data cubes. (5)

(b) What are different Data summarization approaches? Explain analytical characterization of data. (5)

Section -C

Q5 Describe Apriori Algorithm in detail. How can we reduce the complexity of the algorithm? Explain the Support and Confidence metric and its use in Algorithm. (10)

Q6 (a) Describe the Decision tree algorithm. Explain how Gini Index can be used to compare the attributes. (5)

(b) How lazy learners are different. Explain the k-NN algorithm. How the choice of k affects the outcome of algorithm. (5)

Q7 (a) What are the main advantages of Partition around mediods algorithm over k-means algorithm? How it can be further improved? (5)

(b) What are multimedia databases? Explain the Multimedia mining approaches using an example. (5)

x-x-x

Exam. Code 0[illegible]
Sub. Code [illegible]

(Computer Science & Engineering)
Sixth Semester
Elective
CS-60[illegible]: Data Mining [illegible]

Time allowed: 3 Hours Max. Marks: 60

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

Section-A

I. (a) What is data integration? (2[illegible])

(b) Differentiate between [illegible] and d-weights.

(c) What are the measures used to compare the performance of Classification Algorithms?

(d) What is data discretization?

(e) What is difference between time series and sequence data base?

Section-B

II. (a) Highlight the differences between ROLAP, MOLAP and HOLAP. Which [illegible] organisation?

(b) What is multidimensional model of data warehouse? Why full materialization is not [illegible] Explain by [illegible] a suitable example. (5)

III. (a) [illegible] the number of customers visited the store [illegible] day:

[illegible] 200, 250, 320, 300, [illegible], 450, 50, 540, 160, [illegible] 470, [illegible] 120, 4[illegible]

[illegible], 229, 210, 220

Pre-process the data and find first and third quartile of data. Show the [illegible] and quantile plot, divide them [illegible] smooth these by bin [illegible]

(b) What is the purpose of [illegible] Factor? What are the advantages of using these? (5)

IV. (a) How data cubes are computed? Describe the different operations that can be performed on data cubes. (5)

(b) [illegible] dimensional [illegible] Explain [illegible] characterization of data. (5)

Section-C

V. Discuss Apriori algorithm in detail. How can we reduce the complexity of the algorithm? Explain the support and confidence [illegible] in this algorithm. (10)

VI. (a) Discuss [illegible] Decision tree algorithm [illegible] Gini index can be used to compute the difference. (5)

(b) How [illegible] are different. Explain [illegible] K-NN algorithm. How the choice of K affects the outcome of algorithm. (5)

VII. (a) What are the main advantages of [illegible] clustering algorithm over k-means algorithm? How can [illegible] further improved? (5)

(b) What are outliers [illegible] Explain the [illegible] approaches using an example. (5)