Exam.Code:0918
Sub. Code: 6413

2023

B.E. (Computer Science and Engineering)

Sixth Semester

CS-605: 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) What is Data Staging Process?

(b) Illustrate the difference between precision and recall metric.

(c) Highlight the main limitations of Apriori algorithm.

(d) What is training bias? How can we reduce it?

(e) Compare the advantages of Lazy and Eager learners for different applications.

(5x2)

UNIT - I

II. (a) What are the different ways to define the schema of data warehouse? How can we compute the size of data warehouse from its schema?

(b) Why do we perform data aggregation in warehouse? Highlight the main advantages.

(6,4)

III. (a) Describe the role of metadata in data warehouse? How metadata repositories are maintained?

(b) What are the different ways to determine the significance of attributes? How it can be used for data reduction?

(2x5)

IV. (a) What is concept description? Differentiate it from concept characterization and discrimination.

(b) What are different data transformation techniques used during ETL process?(2x5)

P.T.O.

UNIT - II

V. (a) What is the computation complexity of Frequent Pattern mining techniques? Describe the ways to reduce the complexity.

(b) What are different constraints that can be used in association rule mining? (6,4)

VI. (a) What is logistic regression? Illustrate the role of learning parameter.

(b) What are large margin classifiers?

(c) How overfitting in classification handled? (4,4,2)

VII. (a) Describe the clustering algorithm used for large datasets.

(b) Illustrate the process and outcome of spatial data mining? (6,4)

x-x-x