Graphic Era

Mid Term (Odd) Semester Examination October 2024

Roll no.....

Name of the Course and semester: B. Tech CSE 7th semester
Name of the Paper: Data Warehousing and Data Mining
Paper Code: TCS 722
Time: 1.5 hour

Maximum Marks: 50

Note:
(i) Answer all the questions by choosing any one of the sub questions
(ii) Each question carries 10 marks.

Q1. (10 Marks) CO1

a. Explain the concept of data mining. What are the main functionalities of data mining, and how does it provide valuable insights for organizations in various sectors?

OR

b. Discuss the different forms of data preprocessing in data mining. How do data cleaning methods, such as handling missing values, noisy data, and inconsistent data, improve the quality of data for analysis?

Q2. (10 Marks) CO1 & 2

a. Describe the process of data cleaning in detail. How do techniques like binning, clustering, regression, and human inspection help in identifying and correcting noisy data? Provide examples for each method.

OR

b. What is data reduction in the context of data mining? Discuss the various techniques used in data reduction, such as data cube aggregation, dimensionality reduction, data compression, and numerosity reduction, and explain their significance in simplifying large datasets.

Q3. (10 Marks) CO1 &2

a. Analyze the role of clustering, discretization, and concept hierarchy generation in data reduction. How do these methods help in organizing and simplifying complex data structures for efficient data mining?

OR

b. Define concept description in data mining. Explain the processes of data generalization and analytical characterization, and how they contribute to summarizing large datasets.

Mid Term (Odd) Semester Examination October 2024

Q4. (10 Marks) CO2

a. Discuss the analysis of attribute relevance in concept description. How do statistical measures such as central tendency and dispersion help in identifying important attributes in large databases?

OR

b. Explain the use of graph displays in representing basic statistical class descriptions. How do these visualizations assist in understanding the distribution and characteristics of data in large databases?

Q5. (10 Marks) CO2

a. What is association rule mining in data mining? Discuss the Apriori Algorithm and explain how it is used to mine single-dimensional Boolean association rules from transactional databases.

OR

b. Compare the mining of multilevel and multi-dimensional association rules. How are these types of association rules extracted from transaction and relational databases, and what are their applications in data analysis?