Graphic Era

End Term (Odd) Semester Examination December 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: 3 hours   Maximum Marks: 100

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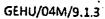   (i) Answer all the questions by choosing *any two of the sub questions* .
   (ii) Each question carries 10 marks.

Q1	(10 marks)	
(a)	Explain the significance of data preprocessing in data mining. How does data cleaning handle missing values, noisy data, and inconsistent data?	CO1
(b)	A dataset contains the following values: [5, 12, 14, 7, 9, 12, ?, 10]. Using mean imputation, calculate the missing value.	
(c)	Compare and contrast different methods of data reduction (e.g., dimensionality reduction, clustering, and numerosity reduction). Propose a scenario where each method would be most effective.	
Q2	(10 marks)	
(a)	Define concept description in data mining. How does it support data generalization and attribute relevance analysis?	CO2
(b)	Using the Apriori algorithm, find all frequent itemsets for a transactional database with a minimum support of 50%:. Transactions: T1: {A, B, C} T2: {A, C, D} T3: {A, B} T4: {B, C} T5: {A, B, C, D}	
(c)	Analyze the differences between single-dimensional and multi-dimensional association rule mining. Discuss their practical applications.	

End Term (Odd) Semester Examination December 2024

Q3	(10 marks)	CO3
(a)	Explain the working of the Decision Tree algorithm. How does it decide the best attribute for splitting at each step?	
(b)	Using the K-Nearest Neighbors algorithm (k=3), classify a data point (2, 3) based on the following training data:. Data: Point 1: (1, 2) → Class A Point 2: (3, 4) → Class B Point 3: (2, 5) → Class A Point 4: (4, 2) → Class B	
(c)	Compare and contrast hierarchical clustering methods (CURE, Chameleon) with density-based methods (DBSCAN, OPTICS). Propose an ideal use case for each method.	

Q4	(10 marks)	CO4
(a)	Differentiate between star schema, snowflake schema, and fact constellation schema in data warehouse modeling.	
(b)	A data warehouse is designed for sales analysis. Construct a multi-dimensional data model with sales as a fact table and dimensions such as time, location, and product. Provide a schematic diagram.	
(c)	Critically evaluate the three-tier architecture of a data warehouse. Design an improved architecture for a hypothetical e-commerce platform.	

End Term (Odd) Semester Examination December 2024

Q5	(10 marks)	
(a)	Discuss the functionalities of MOLAP, ROLAP, and HOLAP servers. How do they address different OLAP requirements?	CO5
(b)	Suppose a company's data warehouse requires weekly backups of 1TB of data. Calculate the time required to back up the warehouse if the transfer rate is 100MB/s.	
(c)	Evaluate the challenges of tuning a data warehouse for performance. Design a strategy to improve the performance of an existing data warehouse.	