



22152-25177-27779-27899

Library
Brainware University
398, Ramkrishnapur Road, Barasat
Kolkata, West Bengal-700125



BRAINWARE UNIVERSITY

Term End Examination 2025-2026

Programme – B.Tech.(CSE)-AIML-2021/B.Tech.(CSE)-DS-2021/B.Tech.(CSE)-AIML-2022/B.Tech.(CSE)-DS-2022/B.Tech.(CSE)-AIML-2023/B.Tech.(CSE)-DS-2023/B.Tech.(CSE)-2023

Course Name – Data Mining and Data Warehousing
Course Code - PCC-CSM602/PCC-CSD602/PCC-CSG602
(Semester VI)

Full Marks : 60

Time : 2:30 Hours

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A

(Multiple Choice Type Question)

1 x 15=15

1. Choose the correct alternative from the following :

- (i) Locate what does the term "dimensional modeling" refer to in data warehousing?
 - a) Modeling 3D graphics for visualization
 - b) Designing databases for online transactions
 - c) Organizing data for analytical purposes
 - d) Processing real-time data streams
- (ii) Identify the purpose of data replication in data warehousing?
 - a) Reducing data redundancy
 - b) Increasing data availability
 - c) Enhancing data security
 - d) Storing historical data
- (iii) Exemplifying Data Warehouse is called as _____ warehouse.
 - a) Read only.
 - b) Write only.
 - c) Read-write only.
 - d) None.
- (iv) The record cannot be updated in _____.
 - a) A..OLTP
 - b) B.Files
 - c) C.RDBMS
 - d) D. Data Warehouse
- (v) Bill Inmon has estimated _____ of the time required to build a data warehouse, is consumed in the conversion process.
 - a) A. 10 percent.
 - b) B. 20 percent.
 - c) C. 40 percent
 - d) D. 80 percent.
- (vi) Identify the classification technique constructs a decision tree based on the attributes of the dataset
 - a) K-Nearest Neighbor
 - b) Support Vector Machine
 - c) Decision Trees
 - d) Naive-Bayes Classifier
- (vii) Determine what is Web Content Mining concerned with?
 - a) Analyzing the layout of a webpage
 - b) Analyzing the structure of a webpage

- c) Extracting useful information from web documents d) Analyzing user behavior on the web
- (viii) Select what is the objective of Web Structure Mining?
- a) To analyze the layout of a webpage b) To analyze the content of a webpage
- c) To analyze the hyperlink structure between web pages d) To analyze the user behavior on the web
- (ix) Identify the objective of clustering algorithms.
- a) Grouping similar data points together b) Predicting future data points
- c) Identifying outliers d) Dimensionality reduction
- (x) Select the key concept behind the K-Means clustering algorithm.
- a) Iteratively minimizing the sum of squared distances b) Maximizing silhouette score
- c) Performing spectral decomposition d) Optimizing entropy
- (xi) Select the process of outlier detection in clustering.
- a) Identify data points deviating from the majority b) Group similar data points together
- c) Optimize cluster centroids d) Assess clustering quality
- (xii) Identify, What does a Bayes Classifier rely on.
- a) Decision boundaries b) Feature independence assumption
- c) Decision trees d) Majority voting
- (xiii) Select, what is the primary purpose of selecting association rule mining?
- a) Clustering similar data points b) Predicting continuous values
- c) Discovering relationships between variables d) Classifying data into predefined categories
- (xiv) Considering a scenario where you need to find association rules in multi-level datasets, Choose from the following, which approach would you select and why?
- a) Apriori algorithm, for its ability to handle large datasets efficiently b) K-means clustering, for its simplicity in implementation
- c) Decision Trees, for their interpretability d) Naive Bayes, for its probabilistic approach
- (xv) The numerical values stored in a data cube are called
- a) Dimensions b) Attributes
- c) Measures d) Keys

Library
Brainware University
 398, Ramkrishnapur Road, Barasat
 Kolkata, West Bengal-700125

Group-B

(Short Answer Type Questions)

3 x 5=15

2. Explain an overview of clustering techniques. (3)
3. Explain how does web usage mining play in improving online user experience. (3)
4. Consider the following transaction database consisting of four transactions. Transaction with ID 100 contains items 1, 3, and 4. Transaction with ID 200 contains items 2, 3, and 5. Transaction with ID 300 contains items 1, 2, 3, and 5. Transaction with ID 400 contains items 2 and 5. The minimum support threshold is 50% and the minimum confidence threshold is 70%. Apply the Apriori algorithm to identify all frequent itemsets and generate the strong association rules that satisfy the given thresholds. (3)
5. Explain Discretization and Concept Hierarchy Generation. (3)
6. Justify the performance of classifiers typically evaluated. (3)

OR

Express the curse of dimensionality, and how does it affect classification. (3)

Group-C

(Long Answer Type Questions)

5 x 6=30

7. Explain Drill-Down and Roll-Up operations in OLAP (5)

8. Justify the role of kernel functions in Support Vector Machines (SVMs) and provide examples of commonly used kernels. (5)
9. Analyze and compare Star, Snowflake, and Galaxy schemas for multidimensional databases with respect to structure, performance, and usage scenarios. (5)
10. Define data cleaning. Express the different techniques for Understanding in handling the missing values. (5)
11. Define Data Warehouses and Explain their types in detail. (5)
12. Problem: Build a Decision Tree classifier for customer churn prediction. Requirements: (5)
Handle missing values Apply entropy-based split Perform cost-complexity pruning Compare accuracy before and after pruning

OR

Create a hybrid ensemble combining: Naive Bayes KNN Use majority voting. (5)

Library
Brainware University
398, Ramkrishnapur Road, Barasat
Kolkata, West Bengal-700125
