BRAINWARE UNIVERSITY

Term End Examination 2024-2025
Programme – M.Tech.(CSE)-AIML-2024
Course Name – Data Warehousing and Data Mining
Course Code - MTA10103
(Semester I)

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) Select which of the following is NOT a component of a data warehouse architecture?

a) Data visualization tools
b) ETL processes
c) Data warehouse server
d) Metadata repository

(ii) Select the purpose of a data mart in data warehousing?

a) Storing raw data
b) Providing a subset of data for a specific business unit
c) Processing online transactions
d) Generating real-time reports

(iii) Predict from the following,Removing duplicate records is a process called _____.

a) Recovery.
b) Data cleaning.
c) Data cleansing.
d) Data pruning.

(iv) Select which one is primarily concerned with Concept hierarchy generation in data mining

a) Reducing data dimensionality
b) Creating a hierarchical structure for categorical data
c) Applying machine learning algorithms
d) Cleaning noisy data

(v) Select which data mining technique is suitable for finding hidden patterns in large datasets?

a) Regression analysis
b) Association rule mining
c) Decision trees
d) K-nearest neighbors

(vi) Select a strategy to handle sparsity issues in association rule mining when dealing with large datasets

a) Increase the minimum support threshold
b) Decrease the minimum support threshold
c) Perform feature engineering to reduce sparsity
d) Apply dimensionality reduction techniques

(vii) Choose from the given option which data mining technique focuses on identifying frequent itemsets in a dataset?

a) Clustering
b) Classification
c) Association rule mining
d) Regression analysis

(viii) Choose from the following what is the primary goal of mining association rules?

a) Clustering similar data points
b) Predicting continuous values
c) Identifying relationships between variables
d) Classifying data into predefined categories

(ix) Select the classification technique that 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

(x) Select the primary goal of classification.

a) Predicting continuous values
b) Sorting data into groups based on similarity
c) Reducing the dimensionality of data
d) Clustering data based on features

(xi) Select the algorithm commonly used for decision tree induction.

a) K-Means
b) C4.5
c) DBSCAN
d) Random Forest

(xii) Select the purpose of feature scaling in classification.

a) To remove irrelevant features
b) To convert categorical features into numerical ones
c) To normalize the range of features
d) To increase the dimensionality of the dataset

(xiii) Select the role of distance metrics in clustering algorithms.

a) Measure dissimilarity between data points
b) Determine optimal number of clusters
c) Optimize cluster centroids
d) Assess clustering quality

(xiv) Choose what is the primary purpose of data preprocessing in KDP?

a) To improve the quality of data
b) To deploy machine learning models
c) To visualize data patterns
d) To generate new data concepts

(xv) Predict from the following why is the Apriori algorithm selected for mining association rules in large databases?

a) It has a low computational complexity
b) It can handle noisy data effectively
c) It efficiently prunes the search space
d) It doesn't require prior knowledge about the data

Group-B

(Short Answer Type Questions) 3 x 5=15

2. Test the performance of classifiers typically evaluated. (3)
3. Write down the main objective of classification. (3)
4. Write about the basics of data mining and its applications. (3)
5. List the Data Warehouse Characteristics in detail (3)
6. Evaluate the role of PCA in clustering. (3)

OR

Select and Describe the specific techniques used in hierarchical clustering. (3)

Group-C

(Long Answer Type Questions) 5 x 6=30

7. Explain the Naive-Bayes Classifier, including its underlying assumptions and how it makes predictions. (5)

8. Explain the concept of Support Vector Machines (SVMs) in classification, including how they work and their advantages. (5)
9. Explain the problem definition in clustering (5)
10. Explain the APRIORI algorithm with an example (5)
11. Write what are the Common OLAP Types. Explain in detail. (5)
12. Compare and contrast agglomerative and divisive methods in hierarchical clustering. (5)

OR

Judge why dimension reduction is important in data mining. (5)

****