**BRAINWARE UNIVERSITY**

Term End Examination 2023-2024
Programme – B.Tech.(CSE)-AIML-2021/B.Tech.(CSE)-DS-2021
Course Name – Data Mining and Data Warehousing
Course Code - PCC-CSM602/PCC-CSD602
( 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)  Choose which method is commonly used for data gathering in the field of Human-Computer Interaction?

a) Observing users in their natural environment.

b) Conducting surveys and questionnaires.

c) Analyzing computer code.

d) None of the above.

(ii)  Discover the primary goal of interaction design.

a) To create aesthetically pleasing interfaces.

b) To design interfaces that are easy and intuitive for users to interact with.

c) To develop complex algorithms for computer interaction.

d) To increase the processing speed of computers.

(iii)  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

(iv)  Determine which technique is used to extract the main topics or themes from a collection of text documents?

a) Clustering

b) Sentiment analysis

c) Topic modeling

d) Named entity recognition

(v)  7. From given list, which of the following is not a common text mining application?

a) Email spam filtering

b) Language translation

c) News categorization

d) Social media analysis

(vi)  Choose which statistical method is commonly used in text classification?

a) Linear regression

b) Principal component analysis

c) Naive Bayes

d) K-nearest neighbors

(vii)  Identify the objective of clustering algorithms.

a) Grouping similar data points together

b) Predicting future data points

c) Identifying outliers

d) Dimensionality reduction

147- 429

(viii) Identify a commonly used algorithm for partitioning data into distinct groups.

a) K-Means

b) Hierarchical Clustering

c) DBSCAN

d) SOM (Self-Organizing Maps)

(ix) Select a technique used for dimensionality reduction in clustering.

a) PCA (Principal Component Analysis)

b) Decision Trees

c) Logistic Regression

d) Random Forest

(x) Identify the primary goal of classification.A) Predicting continuous values B) Sorting data into groups based on similarity

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) Identify the algorithm , commonly used for decision tree induction.

a) K-Means

b) C4.5

c) DBSCAN

d) Random Forest

(xii) Identify ,What does a Bayes Classifier rely on.

a) Decision boundaries

b) Feature independence assumption

c) Decision trees

d) Majority voting

(xiii) Select which of the following steps is NOT part of the data preprocessing stage in KDP?

a) Data Integration

b) Data Reduction

c) Data Mining

d) Discretization

(xiv) Select which technique is used for reducing the complexity of data while preserving its integrity in data reduction?

a) Principal Component Analysis (PCA)

b) Linear Regression

c) Random Forest

d) K-Means Clustering

(xv) 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

## Group-B
### (Short Answer Type Questions)                                3 x 5=15

2. Define Data Warehouses and their types in detail.                (3)
3. List down what are the main types of web mining.                 (3)
4. Describe the problem definition in clustering.                   (3)
5. Discuss main objective of classification.                        (3)
6. Describe the concept of outlier detection.                       (3)

### OR

Explain the PCA algorithm for dimension reduction.                  (3)

## Group-C
### (Long Answer Type Questions)                                 5 x 6=30

7. Explain how does text mining contribute to business intelligence and decision-making.   (5)
8. Discuss role of concept hierarchy generation in a data organization and the steps of     (5)
   concept hierarchy generation.
9. Explain the concept of hierarchical clustering and its two main approaches.              (5)
10. Analyze how can outlier detection be performed using clustering techniques.             (5)
11. List and describe the five primitives for specifying the data mining and warehousing tasks.?  (5)
12. Justify the role of kernel functions in Support Vector Machines (SVMs) and provide       (5)
    examples of commonly used kernels.

### OR

Express the concept of Support Vector Machines (SVMs) in classification, including how    (5)
they work and their advantages.


************************************