18420

BRAINWARE UNIVERSITY

Term End Examination 2025-2026
Programme – M.Tech.(CSE)-DS-2025
Course Name – Data Mining and Pattern Recognition
Course Code - MTD10404
(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) Define the term "data mining".
a) Finding missing values
b) Extracting useful patterns from large data sets
c) Removing outliers
d) Normalizing data

(ii) Identify the first step in data preprocessing.
a) Data transformation
b) Data cleaning
c) Feature selection
d) Data reduction

(iii) Analyze pruning in the context of rule generation.
a) Remove rules with low confidence
b) Combine high lift rules
c) Normalize patterns
d) Add noise to data

(iv) Examine the purpose of using evaluation metrics in data mining?
a) Increase accuracy
b) Monitor performance of mining algorithms
c) Visualize clusters
d) Create databases

(v) Identify the antimonotonic property in pattern mining.
a) Confidence always increases
b) If a set is frequent, all subsets are also frequent
c) Support increases with itemset size
d) Pruning is unnecessary

(vi) Explain the concept of succinctness in constraint-based mining.
a) Use of redundant rules
b) Use of vague thresholds
c) Use of constraints that are easy to check
d) Use of complex graphs

(vii) Analyze the first step in Apriori algorithm?
a) Generate association rules
b) Generate frequent 1-itemsets
c) Prune rules
d) Build decision tree

(viii) Define sequential pattern mining.
a) Mining frequent trees
b) Mining itemsets in order

c) Mining graphs
d) Mining structured data

(ix) Identify an application of trajectory pattern mining.

a) Speech recognition
b) Email filtering
c) GPS route tracking
d) Spam detection

(x) Examine the type of method is k-means clustering?

a) Grid-based
b) Partitioning-based
c) Density-based
d) Model-based

(xi) Propose the clustering algorithm is most suitable for discovering arbitrary-shaped clusters?

a) k-means
b) k-medoids
c) DBSCAN
d) SVM

(xii) Identify an internal validation technique for clustering

a) Silhouette coefficient
b) Rand index
c) F1-score
d) Confusion matrix

(xiii) Choose the best application of LVQ

a) Density estimation
b) Classification tasks
c) Parsing syntax trees
d) Dimensionality reduction

(xiv) Define the primary goal of text mining.

a) Summarizing data
b) Extracting structured information from unstructured text
c) Drawing graphs from data
d) Translating documents

(xv) Identify a tool commonly used for web mining.

a) Hadoop
b) Jupyter
c) BeautifulSoup
d) SPSS

Group-B

(Short Answer Type Questions) 3 x 5=15

2. Explain the distinction between statistical and neural approaches in pattern recognition. (3)
3. Illustrate the flow of the FP-Growth algorithm using a suitable example. (3)
4. Outline key features used in graph matching for pattern recognition. (3)
5. Differentiate between k-means and k-medoids in terms of centroid selection and noise sensitivity. (3)
6. Analyze the role of pattern recognition techniques in speech and video analytics with an example scenario. (3)

OR

Compare phrase mining from text data and traditional keyword extraction in terms of accuracy and context. (3)

Group-C

(Long Answer Type Questions) 5 x 6=30

7. Define the concept of data transformation in the context of data preprocessing. (5)
8. Illustrate the difference between parametric and non-parametric methods in supervised learning with examples. (5)
9. Compare the structural representations used in syntactic pattern recognition with those used in graph-based models. (5)
10. Compare traditional machine learning and deep learning in the context of speech and image pattern recognition. (5)
11. Design an integrated framework combining text mining and pattern recognition techniques to detect online hate speech. (5)
12. Compare the behavior of density-based clustering and grid-based clustering in high-dimensional data. (5)

OR

Analyze the significance of silhouette coefficient in validating clustering output. (5)

****