23DAU636

PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

BSc DEGREE EXAMINATION MAY 2025
(Sixth Semester)

Branch – **COMPUTER SCIENCE WITH DATA ANALYTICS**

MINING OF MASSIVE DATA

Time: Three Hours Maximum: 50 Marks

SECTION-A (5 Marks)
Answer **ALL** questions
ALL questions carry **EQUAL** marks (5 x 1 = 5)

1 Which of the following is not a common type of data mining model?
(i) Classification model (ii) Regression model
(iii) Clustering model (iv) Spread sheet model

2 Which of the following distance measures calculates the straight-line distance between two points in a multi-dimensional space, often considered the most common distance metric for clustering?
(i) Euclidean Distance (ii) Manhattan Distance
(iii) Jaccard Distance (iv) Cosine Similarity

3 Which algorithm is commonly used to count distinct elements in a data stream due to its space-efficient nature and probabilistic approach?
(i) K-means Clustering (ii) Decision Tree
(iii) Naïve Bayes (iv) Flajolet-Martin algorithm

4 Which of the following is the direct application of frequent itemset mining?
(i) Social Network Analysis (ii) Market Basket Analysis
(iii) Outlier Detection (iv) Instruction Detection

5 What does SVD decompose a matrix into?
(i) Two matrices with the same dimensions as the original matrix
(ii) A diagonal matrix with singular values and two orthogonal matrices
(iii) A set of eigenvectors and eigenvalues
(iv) A set of principal components and their variances

SECTION - B (15 Marks)
Answer **ALL** Questions
ALL Questions Carry **EQUAL** Marks (5 x 3 = 15)

6 a Explain Hash functions.
OR
b Describe the algorithms in Map Reducing.

7 a Summarise theory of locality sensitive functions.
OR
b Analyze the methods of high degrees of similarity.

8 a Describe Filtering streams in mining.
OR
b Narrate topic sensitive PageRank.

9 a Analyze Hierarchical clustering.
OR
b Summarise the web-issues in online algorithms on Advertising.

Cont...

10 a Explain partitioning of graphs.

OR

b Compare Eigen values and Eigen vectors of symmetric matrices.

SECTION -C (30 Marks)

Answer **ALL** questions

ALL questions carry **EQUAL** Marks (5 x 6 = 30)

11 a Discuss the statistical limits on data mining.

OR

b Justify theory for Map Reduce.

12 a Examine the applications of set similarity.

OR

b Summarise the applications of LSH.

13 a Analyze Stream data model.

OR

b Discuss about efficient computation of Page Rank.

14 a Discuss Apriori algorithm.

OR

b Differentiate Clustering for streams and parallelism.

15 a Elucidate Social networks as Graphs.

OR

b Analyze Principal Component Analysis.

Z-Z-Z END