

**PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)**

**BSc DEGREE EXAMINATION MAY 2026
(Sixth Semester)**

Branch – COMPUTER SCIENCE

MAJOR ELECTIVE COURSE- II : MINING OF MASSIVE DATA

Time: Three Hours

Maximum: 75 Marks

SECTION-A (10 Marks)

Answer **ALL** questions

ALL questions carry EQUAL marks (10 × 1 = 10)

Module No.	Question No.	Question	K Level	CO
1	1	The importance of a word in a document is commonly measured using (i) PCA (ii) K-means (iii) TF-IDF (iv) PageRank	K1	CO1
	2	Input files for a Map task are viewed as consisting of (i) Records only (ii) Elements (iii) Tables (iv) Rows	K2	CO1
2	3	A k-shingle is (i) A word in a document (ii) A paragraph (iii) A sentence (iv) A substring of length k	K1	CO2
	4	Jaccard distance is defined as (i) Jaccard similarity (ii) 1 + Jaccard similarity (iii) 1 – Jaccard similarity (iv) Square of similarity	K2	CO2
3	5	Which of the following is NOT a stream source? (i) Network traffic logs (ii) IoT sensors (iii) Live video feed (iv) Textbook	K1	CO3
	6	PageRank is used to measure (i) Page size (ii) Page loading speed (iii) Importance of web pages (iv) Number of images	K2	CO3
4	7	In the market-basket model, a basket is also called a (i) Transaction (ii) Record (iii) Cluster (iv) Vector	K1	CO4
	8	The BFR algorithm is designed to (i) Replace k-means completely (ii) Execute k-means on very large datasets (iii) Sort clusters (iv) Reduce dimensions	K2	CO4
5	9	A social network can be represented as (i) Stack (ii) Queue (iii) Graph with nodes and edges (iv) Array	K1	CO5
	10	In Principal-Component Analysis, the dataset is treated as (i) Graph (ii) Tree (iii) Stack (iv) Matrix	K2	CO5

Cont...

SECTION - B (35 Marks)

Answer ALL questions
 ALL questions carry EQUAL Marks (5 × 7 = 35)

Module No.	Question No.	Question	K Level	CO
1	11.a.	Explain Bonferroni's Principle with an example.	K2	CO1
		(OR)		
	11.b.	Illustrate complexity theory for Map reduce.		
2	12.a.	Examine shingling of documents in detail.	K3	CO2
		(OR)		
	12.b.	Explore the Applications of Locality-Sensitive Hashing.		
3	13.a.	Explain the stream data model.	K3	CO3
		(OR)		
	13.b.	Determine an architecture of a Spam Farm and explain it.		
4	14.a.	Explain the market-basket model in detail.	K2	CO4
		(OR)		
	14.b.	Discuss the Hierarchical Clustering in a Euclidean Space.		
5	15.a.	Examine the Partitioning of Graphs in detail.	K4	CO5
		(OR)		
	15.b.	Compare Eigenvalues and Eigenvectors and explain it with an example.		

SECTION -C (30 Marks)

Answer ANY THREE questions
 ALL questions carry EQUAL Marks (3 × 10 = 30)

Module No.	Question No.	Question	K Level	CO
1	16	Explain the Hash functions and indexes with suitable examples.	K2	CO1
2	17	Compute the Similarity-Preserving Summaries of Sets.	K3	CO2
3	18	Depreciate the Counting of Distinct Elements in a Stream with suitable example.	K3	CO3
4	19	How can you handle large dataset in main memory? Explain it in detail.	K2	CO4
5	20	Demonstrate the Clustering of Social-Network Graphs with example.	K3	CO5

Z-Z-Z

END