Register No : _____

THE KAVERY ENGINEERING COLLEGE

(An Autonomous Institution)

M.E DEGREE END SEMESTER THEORY EXAMINATIONS, APRIL-MAY 2025

Second Semester

Computer Science and Engineering

BD4251/PCSE2416- Big Data Mining and Analytics

Regulations – 2021/2024

Duration : 3 Hrs ***Maximum : 100 Marks***

PART – A (ANSWER ALL QUESTIONS) (Marks 10*2=20)

Q.No	*Questions*	*BLT*	*CO*	*Marks*
1.	List out any two important feature extraction from large-scale data.	L2	1	2
2.	If we had to abort and restart the computation every time one component failed, then the computation might never complete successfully. Give any 2 solutions to overcome this problem.	L4	1	2
3.	Define a k-shingle for a document.	L1	2	2
4.	How to Generalize Locality-Sensitive Hashing?	L4	2	2
5.	What are the basic Storage Requirements for the DGIM Algorithm?	L3	3	2
6.	Give an example for Datar-Gionis-Indyk-Motwani Algorithm.	L4	3	2
7.	Differentiate Multistage Algorithm and Multi hash Algorithm.	L4	4	2
8.	If the Distribution of the server approaches a limiting distribution v that satisfies v=Mv. It provided two conditions. What are they?	L5	4	2
9.	How to Initializing K in K-Means Algorithm?	L2	5	2
10.	Define Hash Storage of Word Sets.	L1	5	2

PART – B (ANSWER ALL QUESTIONS) (Marks 5*16 = 80)

Q.No		*Questions*	*BLT*	*CO*	*Marks*
11.	a	Explain with suitable example how “bogus” data was handled by Bonferroni’s Principle.	L2	1	16
		Or			
	b	Explain the Following using Map-Reduce Algorithm. i) Union, Intersection, and Difference by MapReduce ii) Computing Projections by MapReduce iii) Computing Selections by MapReduce	L2	1	16
		Or			
12.	a	Elaborate the Following with suitable example. i) Jaccard Distance ii) Euclidean Distances	L2	2	16
		Or			
	b	Explain Locality-Sensitive Hashing for Documents with suitable examples.	L2	2	16

13.	a	i	With a suitable diagram Explain Data-Stream-Management System.	L2	3	10
		ii	Explain Datar-Gionis-Indyk-Motwani Algorithm and Alon-Matias-Szegedy Algorithm.	L2	3	6
			Or			
			Describe the features of The Count-Distinct Problem and The Flajolet Martin Algorithm with suitable example.	L2	3	16
14.	a	i	Explain Architecture of a Spam Farm in Link spam.	L2	4	10
		ii	How to Formulate Hubbiness and Authority after PageRank was first implemented?	L2	4	6
			Or			
	b	i	Explain Architecture of a Spam Farm in Link spam.	L2	4	8
		ii	Explain Toivonen's Algorithm with suitable example.	L2	4	8
11.	a	i	Explain the Alternative Rules for Controlling Hierarchical Clustering.	L2	5	6
		ii	Explain GRGPF Algorithm in Clustering in Non-Euclidean Spaces.	L2	5	10
			Or			
	b	i	Explain the steps involved in the Content-Based Recommendations.	L2	5	8
		ii	Explain the steps involved in Collaborative Filtering.	L2	5	8