Reg. No.												

Question Paper Code	12311

M.E. / M.Tech - DEGREE EXAMINATIONS, NOV / DEC 2023

First Semester

M.E. - Big Data Analytics

20PBDPC102 - BIG DATA MINING AND ANALYTICS

(Regulations 2020)

Duration: 3 Hours | Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

		Marks, K-Level, CO
1.	What is the need of Machine Learning techniques?	*2,K2,CO1*
2.	Define Feature Extraction.	*2,K1,CO1*
3.	What is Jaccard Distance?	*2,K2,CO2*
4.	What is Generalized Locality-Sensitive Hashing?	*2,K2,CO2*
5.	What is Bloom Filter?	*2,K1,CO3*
6.	List the main characteristics of stream sources.	*2,K2,CO3*
7.	What is the Google Solution to Term Spam?	*2,K2,CO4*
8.	Define Random Surfer Model.	*2,K2,CO4*
9.	Which algorithm is suitable to handle data that is too large to fit in main memory?	*2,K2,CO5*
10.	List the different Clustering Strategies.	*2,K2,CO5*

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

11.	a)	Explain Distributed File Systems in detail.	*13,K2,CO1*
		OR	
	b)	Explain about the Statistical Limits on Data Mining.	*13,K2,CO1*
12.	a)	Explain Locality-Sensitive Hashing for Documents.	*13,K2,CO2*
		OR	
	b)	Explain about Shingling of Documents.	*13,K2,CO2*
13.	a)	Explain Data Stream Model with a neat architecture diagram.	*13,K2,CO3*
		OR	
	b)	Explain the idea behind the Flajolet-Martin Algorithm with suitable example.	*13,K2,CO3*

14. a) Explain A-Priori Algorithm with suitable example. *13,K2,CO4*

OR

b) Explain PageRank in detail. *13,K2,CO4*

15. a) Explain K-means Algorithms and also explain how will initialize Clusters for K-Means. *13,K2,CO5*

OR

b) Explain the CURE Algorithm in detail. *13,K2,CO5*

PART - C (1 × 15 = 15 Marks)

16. a) Why did Netflix want/need to do a big data project? *15,K3,CO5*

OR

b) How Facebook recommend the person which has the highest possibility to become friend? Explain. *15,K3,CO5*