PRN:		Course Code :	P70PC103

First Year **M. Tech (AIDS)** End Semester Examination Winter 2024.

Scalable Systems for Data Science 000436

Day and Date: Wed - 4 DEC 2024 Max Marks: **50**

Time : 10:30AM To 12:30PM Duration: **02** hrs

Instructions:

1. All questions are compulsory.
2. Assume suitable data wherever necessary and mention it clearly.

			Marks	Bloom's Level	CO
Q.1		**Attempt the following Question.**			
	a)	Define the Data Mining and Data Warehousing.	2	1	1
	b)	Define a link spam?	2	1	2
	c)	Define Market Basket Analysis.	2	1	3
	d)	Define Hierarchical Clustering?	2	1	4
	e)	Define a Recommender System?	2	1	5
Q.2		**Attempt the following Question.**			
	a)	Write and explain K-means clustering algorithm.	5	2	2
	b)	Draw and explain architecture of Mapreduce.	5	2	2
		OR			
	c)	Explain different data stream clustering methods.	5	2	2
Q.3		**Attempt the following Question.**			
	a)	Write and explain Advantages and disadvantages of HDFS Architecture.	5	2	3
	b)	Explain Data Handling in Data science.	5	2	3
		OR			
	c)	Explain benefits and applications of Market Basket Analysis in detail.	5	2	3
Q.4		**Attempt the following Question.**			
	a)	Explain types of Market Basket Analysis.	5	2	4
	b)	Explain types of Hierarchical Clustering.	5	2	4
		OR			
	c)	Describe real-world applications of Distributed Machine learning.	5	3	4
Q.5		**Attempt the following Question.**			
	a)	Explain Types of Recommender Systems.	5	3	5
	b)	Describe different Challenges for Distributed Machine Learning.	5	2	5
		OR			
	c)	Give Types of collaborative filtering.	5	2	5

ESE