Register Number:……..

KUMARAGURU
college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 2023

(Regulation 2018)

Fifth Semester

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

U18AIE5005 : Mining Bigdata

COURSE OUTCOMES

CO1: Choose B tools to carry out exploratory data analysis and produce effective visualization of given data.

CO2: Perform parallel data processing and duplication with Hadoop and Map-Reduce.

CO3: Identify suitable data model and algorithms for mining mass data set.

CO4: Apply link analysis & mining social network graphs in real time problem.

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:

PART A (10 x 2 = 20 Marks)

(Answer not more than 40 words)

1.	List the advantages of statistical inference.	CO1	[K_1]
2.	Identify the few basic tools for Exploratory Data Analysis.	CO1	[K_3]
3.	Compare and contrast scatter plot and box plot.	CO1	[K_2]
4.	Identify how Fault Tolerance is involved in HDFS.	CO2	[K_3]
5.	Differentiate conventional systems and intelligent data analysis.	CO2	[K_2]
6.	Explain stream computing with an example.	CO3	[K_2]
7.	List few Real-Time Analytics Platform (RTAP) applications.	CO3	[K_1]
8.	What do you mean by Trust Rank?	CO4	[K_1]
9.	Differentiate between page rank and spam mass.	CO4	[K_2]
10.	How is combating link spam implemented?	CO4	[K_1]

Answer any FIVE Questions:

PART B (5 x 16 = 80 Marks)

(Answer not more than 400 words)

11.	Explain the importance of probability distributions with relevant example.	16	CO1	[K_2]
12.	Analyze how the following phases in Map Reduce are interlinked with one	16	CO2	[K_4]

another.

- Shuffle and sort
- Reducer
- Output Phase

13.	a)	What do you mean by filtering streams and why it is important?	8	CO3	[K_1]
	b)	Illustrate decaying window with an example.	8	CO3	[K_2]
14.		Explain how combiners are used to consolidate the Result Vector.	16	CO4	[K_2]
15.		Consider a file of size 400 MB to be load in the cloud, Analyze which file system either GFS or HDFS is suitable to store the file in optimal way. Justify the reason by demonstrating partition of chunk file.	16	CO2	[K_4]
16.		Identify a scenario in which the aspects of Analysis is applied and Reporting is done for the same.	16	CO1	[K_3]

****