Question Paper

MANIPAL ACADEMY OF HIGHER EDUCATION

SIXTH SEMESTER B.TECH AIML END SEMESTER EXAMINATIONS, MAY 2024

BIG DATA ANALYTICS [CSE 3272]

Marks: 50   **Duration: 180 mins.**

Answer all the questions.

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

1)	A)	What are components of Hadoop architecture? Explain each of them.	(5)
	B)	Compare different Hadoop Installation Modes.	(3)
	C)	Summarize characteristics of HDFS.	(2)
2)	A)	With the help of neat diagram explain components of Spark application.	(5)
	B)	Identify ALS algorithm hyperparameters and explain them.	(3)
	C)	Write a program in PySpark to count the number of documents present in a folder.	(2)
3)	A)	What is anomaly detection in a network? Write a program to detect the anomaly in a network.	(5)
	B)	Write the code snippet for one hot encoding.	(3)
	C)	Compare Reduce function with ReduceByKey function.	(2)
4)	A)	What is sliding window? Compare it with State operations. Draw a diagram for sliding window with window length of 2 and slide interval of 2. Write codes for sliding window.	(5)
	B)	What are the drawbacks of conventional summary statistics in analyzing the co-occurrence networks? How do you overcome them?	(3)
	C)	What are the design goals of Spark Streaming?	(2)
5)	A)	What is a state operation? Explain the function responsible for achieving state operation Illustrate state operation with the help of code sample.	(5)
	B)	Compare top and take functions.	(3)
	C)	Write the code for random forest.	(2)

-----End-----