USN | | | | | | | | | | | **17CS82**

Eighth Semester B.E. Degree Examination, July/August 2022
Big Data Analytics

Time: 3 hrs.   Max. Marks: 100

Note: *Answer any FIVE full questions, choosing ONE full question from each module.*

Module-1
1 a. Discuss the various system roles in an HDFS components or deployment.   (10 Marks)
   b. Describe HDFS block replication with an example.   (10 Marks)

OR
2 a. Briefly explain HDFS NameNode Federation, NFS gateway, Snapshots.   (10 Marks)
   b. Write a program to Read and Write HDFS file using java.   (10 Marks)

Module-2
3 a. Discuss the usage of Apache Pig.   (08 Marks)
   b. Explain Apache Sqoop to Acquire Relational data with an example.   (08 Marks)
   c. Give the Apache Flume to acquire data streams.   (04 Marks)

OR
4 a. Demonstrate the working of Hive with Hadoop.   (08 Marks)
   b. Explain YARN application framework with an example.   (08 Marks)
   c. Explain briefly how to manage Hadoop with Apache Ambari.   (04 Marks)

Module-3
5 a. List and explain any 3 areas of applications of Business Intelligence (BI).   (10 Marks)
   b. Define Data Warehouse. Explain design consideration for data warehouse.   (10 Marks)

OR
6 a. What is Data Mining? What are supervised and unsupervised learning techniques? (10 Marks)
   b. What is Data visualization? Explain how visualization tools are used.   (10 Marks)

Module-4
7 a. Using the Data given in Dataset as shown below, create a regression model to predict the
       Test2 from Test1 score. Then predict the score for the one who got a 46 in Test1. (10 Marks)

Test 1	59	52	44	51	42	42	41	45	27	63	54	44	50	47
Test 2	56	63	55	50	66	48	58	36	13	50	81	56	64	50

   b. Write the different steps involved in developing an artificial neural network. (10 Marks)

OR

8 a. Construct a Decision tree that helps to make decision about approving the play of an outdoor game.

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	Normal	True	?

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

(10 Marks)

b. Apply Apriori Algorithm for the following table, assume support count (minsup) = 2.

T_{ID}	items
T_1	I_1, I_2, I_5
T_2	I_2, I_4
T_3	I_2, I_3
T_4	I_1, I_2, I_4
T_5	I_1, I_3
T_6	I_2, I_3

(10 Marks)

Module-5

9 a. List and explain different types of Text Mining applications.   **(10 Marks)**

b. What is Naïve-Bayes technique? Explain its model.   **(10 Marks)**

OR

10 a. What is SVM? With a neat diagram, explain support vector machine model.   **(10 Marks)**

b. Define social network analysis? Explain different types of network topologies.   **(10 Marks)**

* * * * *