62477

Reg. No. P 18 I W 21 5000 5

III Semester M.C.A Degree Examination, June/July - 2023

COMPUTER SCIENCE

Big Data and Analytics (Elective)

(CBCS Y2K20 Scheme)

Paper : 3MCAE2

Time : 3 Hours **Maximum Marks : 70**

Instructions to Candidates:

1 Part-A: Answer any **five** questions. Each carries **6** marks.

2. Part- B: Answer any **four** questions. Each carries **10** marks.

PART-A

Answer any **Five** questions. **Each** question carries **Six** marks. **(5×6=30)**

1. What are the characteristics of Big Data? Briefly explain the challenges in handling Big data.
2. What is the relevance of Data Reduction? Discuss briefly the various Data Reduction techniques.
3. What is Concept hierarchy? What is its significance?
4. Explain classification by Backpropagation.
5. What do you understand by web mining and Text Mining? Discuss each one briefly.
6. Explain the anatomy of the File Read and File Write operations in HDFS.
7. Illustrate and explain any six relational operators in PIG/ PIG LATIN.
8. Compare supervised Learning v/s unsupervised learning.

PART - B

Answer any **Four** questions. Each question carries **Ten** marks. **(4×10=40)**

9. a) Compare structured, semi-structured and unstructured data. **(5)**

   b) What is Data preprocessing? Why is Data preprocessing necessary? **(5)**

[P.T.O.

10. a) Explain the Decision tree classifier. (6)

    b) Briefly discuss the issues of Classification and Prediction. (4)

11. a) Explain Hadoop ecosystem and its components. (8)

    b) Compare SQL V/s No SQL databases (2)

12. Write a java program using mapreduce to perform word count. (10)

13. What is HIVE? With illustrations explain the HIVE DDL and DML commands. (10)

14. Write short notes on:

    a) HBASE (5+5)

    b) Data Analytics with R
