

S. B. Roll. No.....

BIG DATA
5th Exam/CSE/IT/3195/Nov'24
(For 2018 Batch Onwards)

Duration: 3Hrs.

M.Marks:75

SECTION A

Q1. Do as directed.

15x1=15

- a. What is Term Frequency?
- b. Briefly explain Big Data analytics.
- c. Explain RDD in brief.
- d. What is HADOOP?
- e. What is Machine Learning?
- f. How many forms are there of Big Data?
- g. What are advantages of NoSQL database?
- h. R made its first appearance in which year?
- i. What is HDFS?
- j. What is Data Mining?
- k. Any two roles in data Science Progress.
- l. Full form of HPCC.
- m. What is R language?
- n. Which platform is used to run HADOOP?
- o. What is extension of a R file?

SECTION-B

Q2. Attempt any six questions.

6x5=30

- i. Explain any 5 big data tools.
- ii. What are five V's of big data?
- iii. Explain how sampling is done for modelling and validation.
- iv. What is Data Serialization?
- v. What are various challenges in Big Data?
- vi. Explain working of MAP Reduce in HADOOP?
- vii. Explain briefly NoSQL.
- viii. List any 5 features of HADOOP.

SECTION-C

Q3. Attempt any three questions.

3x10=30

- a. Discuss HADOOP structure in detail.
- b. Explain various statistical models in R.
- c. What is Map Reduce? Also explain various stages of it.
- d. Explain R tool and how data is processed using R tool?
- e. Write short note on
i) Mongo DB ii) Hbase