

Total No. of Questions : 8]

[Total No. of Printed Pages : 2

Roll No

IT-8003(2)-CBGS

B.E. VIII Semester

Examination, December 2020

Choice Based Grading System (CBGS)

Data Science

Time : Three Hours

Maximum Marks : 70

- Note:** i) Attempt any five questions.
ii) All questions carry equal marks.

1. Explain the following:
 - i) Naive bayes classifier
 - ii) Evolution of data management
2. a) Define Big data. Explain Big data management architecture.
b) Define Structured Data. Explain sources of big structured data.
3. a) Explain Linear regression with Non linear substitution.
b) Explain Linear Correlation with suitable example.
4. a) Write the differences between Parallel vs Distributed Processing.
b) Explain Life cycle of Data analytics.

IT-8003(2)-CBGS

PTO

[2]

5. a) Explain the architecture of Hive and how it is useful for data analysis.
b) What is Hadoop? Explain its infrastructure and HDFS concept.
6. a) Explain data manipulation and statistical analysis with R.
b) Explain data types used in R and loops and conditional statements.
7. a) Show how would you use the sampling distribution and re-sampling
b) Explain core components of analytical data architecture.
8. Write short notes on any two:
 - i) Unstructured Data
 - ii) K-Means clustering
 - iii) Arrays and Matrices

IT-8003(2)-CBGS