

Roll No

IT-8003 (2) (CBGS)

B.E. VIII Semester

Examination, May 2019

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. a) What is Big data? Describe the main features of a data analytical system.
b) What is Data Science? Also, list the differences between supervised and unsupervised learning.
2. a) Describe in detail about the role of statistical models in Big Data.
b) What do you understand by the term Normal Distribution?
3. a) Explain the architecture of Hive and how it is useful for data analysis.
b) What is Hadoop? Explain its infrastructure and HDFS concept.
4. a) Explain Naive Bays classifier with example.
b) Differentiate Hadoop vs distributed database.

5. a) Show how would you use the sampling distribution and re-sampling
b) Explain core components of analytical data architecture.
6. a) Explain data manipulation and statistical analysis with R.
b) Explain data types used in R and loops and conditional statements. http://www.rgpvonline.com
7. a) What is parallel and distributed processing? Explain with example.
b) Explain the basic building blocks of Hadoop with a neat sketch.
8. Write short notes
 - i) Structured and Unstructured Data
 - ii) K-Means clustering
 - iii) Arrays and Matrices
