Roll No ....

IT-702(A)-CBGS
B.Tech., VII 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. a) How are Big Data and Data science related?
    b) Describe Data Analytics life cycle.

2. a) Define regression. Give the types of regression.
    b) How are association rules mined from large databases?

3. a) How to calculate the accuracy of a binary classification algorithm using its confusion matrix?
    b) Differentiate Agglomerative and Divisive Hierarchical Clustering?

4. a) What are the differences between regular File System and HDFS?
    b) What benefits did YARN bring in Hadoop 2.0 and how did it solve the issues of MapReduce v1?

5. a) If you have an input file of 350 MB, how many input splits would HDFS create and what would be the size of each input split?
    b) What role do RecordReader, Combiner, and Partitioner play in a MapReduce operation?

6. a) What are the components used in Hive query processors?

   b) What are the key differences between Hive and Pig?

7. a) What are the key components of HBase?

   b) How is Sqoop different from Flume?

8. a) Compare R and Python programming languages for Predictive Modelling.

   b) What is the best way to communicate the results of data analysis using R language?

******

IT-702(A)-CBGS