H

Roll No. .....

TCS–702

B. TECH. (CSE)
(SEVENTH SEMESTER)
END SEMESTER EXAMINATION, 2021-22
BIG DATA ANALYTICS

Time : Three Hours

Maximum Marks : 100

Note : (i) All questions are compulsory.

(ii) Answer any *two* sub-questions among **(a), (b) and (c)** in each main question.

(iii) Total marks in each main question are **twenty.**

(iv) Each question carries 10 marks.

1. (a) What do you mean by big data ? What is data life cycle ? Explain all the phases. (CO1)

(b) Explain the technological components in Big data. Mention some industry which contribute in the big data. Also explain how this technique help in growth. (CO1)

(c) What do you understand by data pre-processing, preparation and data cleaning ? Explain their importance. (CO1)

2. (a) What is Predictive Modeling ? Also describe the techniques of it. (CO2)

(b) What do you mean by the term Data Visualization ? Explain Box and Whisker plot. (CO2)

(c) What is a vector( ) in R ? Explain exploratory data analysis in terms of R. (CO2)

3. (a) Write about K-means clustering and Naïve Bayes algorithm. (CO3)

(b) Explain decision tree, classification and regression. (CO3)

(c) What is the difference between linear and logistics regression ? (CO3)

4. (a) What do you mean by Hadoop ? Explain Hadoop Ecosystem. With reference, explain MapReduce. (CO4)

(b) Explain data node, name node, secondary namenode, block in HDFS. (CO4)

(c) What are the deamons of MapReduce ? Explain them briefly. (CO4)

5. (a) Explain for three deliverables of a big data life cycle project. (CO5)

(b) Define PIG and HIVE.Spark and Flume. (CO5)

(c) The difference in Parallel and Distributed Computing. Explain. (CO5)

420