Chennai. It has collected data from around 300 sportsmen, what sports they play and what are their favorite sports drinks. Since the data is fairly small, it is stored in an excel sheet. As a big data scientist you are entrusted with researching the question : What are the three most popular sports drinks in the four metros of India (Delhi, Chennai, Mumbai and Kolkata) ? Devise a strategy in terms of the five steps of data science.　　　**10 Marks (CO2 and CO3)**

H　　　　**Roll No.** ....

TCS–671

B. TECH. (CSE)
(SIXTH SEMESTER)
MID SEMESTER EXAMINATION,
April/May, 2022

BIGDATA STORAGE AND PROCESSING

Time : $1\frac{1}{2}$ Hours

Maximum Marks : 50

Note : (i) Answer all the questions by choosing any *one* of the sub-questions.

(ii) Each sub-question carries 10 marks.

1. (a) Summarize the characteristics of Bigdata giving real life examples. **10 Marks (CO1)**

OR

(b) It is said that almost 70% work of any Bigdata project is the (Pre) processing step. Summarize the different types of tasks that are done in this step.

10 Marks (CO1)

2. (a) Differentiate between Statistical Modeling and Algorithmic Modeling in detail.

10 Marks (CO2)

OR

(b) For the class Wolves draw the Confusion Matrix. Then calculate the Precision, Recall and F1 score. 10 Marks (CO2)

	FOX	JACKAL	WOLF
FOX	1200	400	190
JACKAL	400	500	450
WOLF	210	350	1300

3. (a) List the different components of the Hadoop Ecosystem. Explain the HDFS component in detail.

10 Marks (CO1 and CO2)

OR

(b) Is Apache Spark a part of Hadoop ecosystem ? Explain the Spark architecture in detail. 10 Marks (CO1 and CO2)

4. (a) What is Name Node and Data Node ? Compile the roles of Name Node and Data Node in Hadoop. 10 Marks (CO2)

OR

(b) Provide a comprehensive description of the following components of Hadoop : YARN, MapReduce, HBASE

10 Marks (CO2)

5. (a) Say you are building a Recommender system to recommend some product. Construct the Oozie workflow for the same. 10 Marks (CO2 and CO3)

OR

(b) Say you have recently joined an organization based in New Delhi that distributes sports drinks as a bigdata scientist. The organization is 5 years old, it has two branch offices in Mumbai and