

CDOE
MASTER OF COMPUTER APPLICATIONS (CBCS-2019 COURSE)
M.C.A. SEM - V : WINTER :- 2021
SUBJECT: DATA SCIENCE

Day : Tuesday
Date 22-02-2022

W-22260-2021

Time : 10:00 AM-01:00 PM
Max. Marks: 60

N.B.

- 1) Q.No. 4 from Section I is **COMPULSORY**.
 - 2) Answer **ANY TWO** questions from Q.1,2,3 in Section – I.
 - 3) Answer **ANY TWO** questions from Q.5,6,7 in Section – II.
 - 4) All questions carry **EQUAL** marks.
 - 5) Answer to both the sections should be written in **SAME** answer book.
 - 6) Draw labeled diagrams **WHEREVER** necessary.
-

SECTION – I

- Q.1** What is Data Science? Explain its relation to Data Mining in detail. (12)
- Q.2** List different visualization techniques. Explain boxplot and scatterplot in detail. (12)
- Q.3** What is exploratory data mining? List its methods and explain any one. (12)
- Q.4** Write short notes on **ANY TWO** of the following: (12)
- a) Hypothesis testing
 - b) Measures of performance
 - c) Techniques of data split
 - d) Hadoop ecosystem

SECTION – II

- Q.5** What is data discretization? Distinguish between unsupervised and supervised discretization. (12)
- Q.6** What is clustering? Explain Hierarchical clustering in details. (12)
- Q.7** A dataset contains list of cricket players and their scores in last five innings. Define and describe corresponding dataset and write statements for followings. (12)
- i) Read data from dataset.
 - ii) Identify highest run scorer.
 - iii) Identify valuable player according to average runs score.
 - iv) Compare the performance of any two players through visualization technique.
