VIT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act 1956)

School of Computer Science and Engineering

Fall Semester 2023-2024

Continuous Assessment Test – 1

SLOT: E1

Programme Name & Branch : MID

Course Name & code: Data Science Programming (CSI3004)

Class Number (s): VL2023240104182 / VL2023240104180

Faculty Name (s): Dr. Sayan Sikder

Exam Duration: 90 Mins.   Maximum Marks: 50

Q.No.	Question	Max Marks
1.	Define Data Science and explain the six steps of the process. Inter-relate Mathematics, Computer Science and Domain expertise with the help of a Venn diagram.	10
2.	Contrast between supervised and unsupervised learning. Implement a K means clustering Algorithm for the following data set where K=2. Name / Matches / Goals Ronaldo / 10 / 8 Roberto Carlos / 7 / 1 Pepe / 9 / 1 Zidane / 10 / 9 Rivaldo / 8 / 6 Maldini / 11 / 2	10
3.	a) Define Digital Universe and Information Commons. (6 Marks) b) What are the phases of data science project life cycle. Elaborate the possibilities and practices of the data preparation phase with examples. (4 marks)	10
4.	a) Define Conditional Probability. Two dice are rolled, if it is known that atleast one of the dice always shows 4, find the probability that the numbers appeared on the dice have a sum 8. (6 Marks) b) What are the challenges in Data Science Technology (4 Marks)	10
5.	a) Comment on the K in KNN algorithm with an appropriate example. (4 Marks) b) Write a short note on decision trees explaining the important components with an comprehensive example (6 Marks)	10

The table for Q2:

Name	Matches	Goals
Ronaldo	10	8
Roberto Carlos	7	1
Pepe	9	1
Zidane	10	9
Rivaldo	8	6
Maldini	11	2