Code No: R194105B **R19** Set No. 1

IV B.Tech I Semester Regular Examinations, November – 2022
DATA SCIENCE
(Computer Science and Engineering)

Time: 3 hours   **Max. Marks: 75**

Answer any FIVE Questions
ONE Question from Each unit
All Questions Carry Equal Marks
*****

UNIT-I

1 a) Differentiate **Tuple** and **List** in python. Focus on various methods used in both. [7]
   b) Develop a python code to read n values and sort them. [8]
   (OR)
2 a) Write a Python program to print the sum of factors of a given number [7]
   b) Define module. Give an example for module in python [8]

UNIT-II

3 a) Write about Simpson's paradox. Give its implications in data. [7]
   b) Develop a code to plot the curve $y = x^2+3$ using *matplotlib* package [8]
   (OR)
4 a) How to define a vector and a matrix in python? Compare both data structures. [7]
   b) What is the use of scatter plot? Give an example code for scatter plot and explain it. [8]

UNIT-III

5 a) Define Bayes's theorem. Give its applications [7]
   b) Give the significance of cleaning the data and dimensionality reduction before working on data. [8]
   (OR)
6 a) Write a python code to read a file and print it on console. [7]
   b) Explain central limit theorem. Give its applications. [8]

UNIT-IV

7 a) Explain the concept of machine learning. Summarize about the package used for it in python. [7]
   b) Differentiate linear regression and logistic regression [8]

(OR)

8 a) Differentiate Bias and Variance with suitable example.   [7]
   b) Explain the k – nearest neighbours algorithm   [8]

UNIT-V

9 a) Explain the idea and model of clustering focus on selection of k in k [7]
       means clustering.
   b) Give the important features of Numpy and pandas libraries.   [8]

(OR)

10 a) Explain about Bottom-up Hierarchical Clustering   [7]
   b) Differentiate user based collaborative filtering with item based   [8]
       collaborative filtering