H.T.No. | | | | | | | | | | |

Code No: CT3545 **SRGEC-R20**

III B.Tech II Semester Regular Examinations, May 2023

DATA SCIENCE

(Artificial Intelligence and Data Science)

Time: 3 Hours **Max. Marks: 70**

Note: Answer one question from each unit.
All questions carry equal marks.

5 × 14 = 70M

UNIT-I

1. a) Develop a Python program for implementing KNN. (7M)
   b) Explain different descriptive statistics measures. (7M)

(OR)

2. a) Distinguish supervised learning and unsupervised learning. (7M)
   b) Develop a Python program for calculating simple linear regression. (7M)

UNIT-II

3. a) Explain feature extraction and selection methods. (7M)
   b) Explain the difference between classification and regression. (7M)

(OR)

4. a) Diagnose the causes for producing a model that performs well on the data you trained it on but generalizes poorly to any new data. (7M)
   b) State the Bayes theorem and explain the Naïve Bayes classifier. (7M)

UNIT-III

5. a) How draw a scatter plot with labels, legends and title in Matplotlib? (7M)
   b) Explain two-sample t-test with equal variance. (7M)

(OR)

6. a) Describe the approach used to perform paired t and u test. (7M)
   b) Write the formulae to calculate the quartiles and discuss with an example. (7M)

UNIT-IV

7. a) Illustrate the process of imputing and tracking missing values. (7M)
   b) Develop a Python code to retrieve data from CSV file format. (7M)

(OR)

8. a) Describe the key steps in data wrangling process. (7M)
   b) Write a Python code to retrieve data from JSON file format. (7M)

UNIT-V

9. Compute the eigenvalues and eigenvectors of the matrix A. A<- matrix(c(13, -4, 2, -4, 11, -2, 2, -2, 8), 3, 3, byrow=TRUE). (14M)

(OR)

10. How to implement the mechanism of data de-duplication? Discuss with suitable examples. (14M)