Code No: 156GA **R18**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech III Year II Semester Examinations, August/September - 2024
FUNDAMENTALS OF DATA SCIENCE
(Computer Science and Engineering – Artificial Intelligence and Machine Learning)

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

Note: i) Question paper consists of Part A, Part B.
ii) Part A is compulsory, which carries 25 marks. In Part A, Answer all questions.
iii) In Part B, Answer any one question from each unit. Each question carries 10 marks and may have a, b as sub questions.

PART – A

(25 Marks)

1.a)	Define overfitting.	[2]
b)	What are the roles exist in data science?	[3]
c)	What is an attribute? How to describe an attribute?	[2]
d)	Define range, variance and standard deviation.	[3]
e)	What is a data frame? Give an example.	[2]
f)	How is a vector differs from a list?	[3]
g)	Define Nested functions in R.	[2]
h)	Discuss briefly about looping over list.	[3]
i)	Define data cube.	[2]
j)	What is regression and write its applications.	[3]

PART – B

(50 Marks)

2.a) Explain the steps to install RStudio in Linux.
b) Write the applications of data science. [5+5]

OR

3.a) Describe the different tools for data scientists to analyze data.
b) What is R programming language and write its features and explain various data types in R. [5+5]

4.a) Explain various ways to measure central tendency in data.
b) Give a note on asymmetric attributes, nominal attributes and ordinal attributes. [5+5]

OR

5.a) Identify various probability distributions commonly used for statistical modeling.
b) With example, explain the differences between discrete and continuous attributes. [5+5]

6.a) Discuss different levels of factor.
b) What is a Named list? Explain steps to create named list and access its elements. [4+6]

OR

7.a) Write a program to sort data frames.
b) Explain various properties and operations on matrix. [4+6]

8.a) Describe relational and logical operators in “R” programming.
b) Explain different conditional statements in “R” programming. [5+5]

OR

9.a) Elaborate mathematical functions in “R” programming.
b) Explain different loop statements for data frames and illustrate with an example. [5+5]

10.a) Discuss icon based visual technique.
b) Write PCA algorithm and steps for data reduction. [5+5]

OR

11.a) Give a note on data reduction strategies.
b) Explain attribute subset selection method for data reduction. [5+5]

---ooOoo---