ACE
Engineering College
(An Autonomous Institution)

Question Paper Code: CM512PE/ CD502PC

ACE-R20

Semester End Examination
III B. Tech- I Semester Regular- JAN/FEB -2023
INTRODUCTION TO DATA SCIENCE
Common to CSM, CSD

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

H. T. No										

Note: i) This question paper contains two parts A and B.
ii)Part A is compulsory which carries 20 marks. Answer all questions in Part A.
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 MARKS: 10*2=20

Q.No: 1	Question	Marks
a)	Differentiate populations and samples in statistics.	2
b)	What is overfitting? Is it good to fit any model?	2
c)	What is a data object. List the data objects related to sales database.	2
d)	What is an outlier? Write one advantage of the outliers.	2
e)	Create a vector containing the values 0, 0.25, 0.5, 0.75, and 1?	2
f)	What is a factor in R programming?	2
g)	Write the output for the following code: y = c("a", "bc", "def") y == "a" x = c(1, 7, 1, 20) which(x > 2)	2
h)	Write any 4 math functions in R.	2
i)	Write the advantage of data reduction.	2
j)	Define discrete wavelet transform (DWT).	2

PART- B MARKS: 5*10=50

Q.No	Question Description	Marks
2.	Explain the basic data types used in R programming in detail.	10
	(OR)	
3	Explain the current landscape perspectives related to Big Data and Data Science.	10
4	Define attribute. Explain different types attributes with suitable examples.	10
	(OR)	
5.	Discuss about Boxplot, Scatter plot and histograms.	10
6	What is a data frame? Explain about extending and sorting data frames with example.	10
	(OR)	
7	What is a list? Explain different operations on lists with suitable examples.	10
8	Describe binary search implementation using recursion.	10
	(OR)	
9	Discuss about loops in R programming with examples.	10
10	Explain greedy methods for attribute subset selection.	10
	(OR)	
11	Briefly explain Geometric projection and Icon based visualization techniques.	10