Code No: 155EQ **R18**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech III Year I Semester Examinations, January/February - 2023
DATA SCIENCE
(Common to CSBS, CSIT)

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)	What is big data?	[2]
b)	What is the need for data science?	[3]
c)	Define data analysis.	[2]
d)	Write a short note on KNN.	[3]
e)	What is the importance of machine learning?	[2]
f)	List the applications of machine learning.	[3]
g)	What is the main motivation of random forest algorithm?	[2]
h)	What is the importance of data wrangling?	[3]
i)	List out the benefits with data visualization.	[2]
j)	What are the key principles of data visualization?	[3]

PART – B

(50 Marks)

2.a) Describe the skill sets needed for data scientists.
b) Compare data science with big data. [5+5]

OR

3. Explain the following terms.
a) Probability distribution.
b) Datafication. [5+5]

4. Write KNN algorithm and illustrate with an example. [10]

OR

5.a) Write short notes on philosophy of EDA.
b) Explain data science process in detail. [5+5]

6.a) "Linear regression algorithm is a poor choice for filtering spam". Justify this statement.
b) With respective to filtering spam, outline KNN algorithm. [5+5]

OR

7. Write Naive Bayes algorithm and illustrate with an example. [10]

8. Explain the following terms:
   a) Feature selection with decision tree method
   b) APIs and tools for scrapping the web. [5+5]

OR

9.a) Describe feature selection with random forest algorithm.
   b) Compare feature selection with feature extraction. [5+5]

10.a) "Multimedia is the most popular source of big data". Justify this statement.
   b) Describe the best data visualization techniques for small and large data. [5+5]

OR

11. Present a detailed summary on the next generation of data scientist. [10]

---ooOoo---