Code No.: B41FD3868 **Date: 03.01.2022**

CVR COLLEGE OF ENGINEERING

UGC Autonomous Institution **-** Affiliated to JNTUH

R18

B. Tech IV Year I Sem. Main *Exams Jan - 2022* (2018 Batch)
Subject: Fundamentals of Data Science
(OE)

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

Note:
1. Please verify the regulation of question paper and subject name
2. Question Paper Consists of Part-A and Part B
3. Assume required data, if not given in the question

PART – A **(10x2= 20 Marks)**
(Answer ALL Questions)

1.	Are Big Data and Data science are different or same Explain?	BL1(CO1)
2.	What is datafication?	BL2(CO1)
3.	What are the tools of Data Visualization?	BL1(CO2)
4.	Discuss the role of Descriptive Statistics in EDA.	BL2(CO2)
5.	Discuss how Regression helpful in data sciences Problems?	BL1(CO3)
6.	Why Naive Bayes is used in Data Sciences explain?	BL4(CO3)
7.	Explain the role of SVD discuss why this is helpful in data handling?	BL1(CO4)
8.	What do you mean by data product? Discuss with an appropriate example?	BL2 (CO4)
9.	How to study the impact of the input Features on the output in data set?	BL3(CO5)
10.	Describe Feature Selection versus Feature Reduction?	BL4(CO5)

PART – B **(5x10 = 50 Marks)**
(Answer ALL Questions)

11. Explain the method of fitting Poission distribution with an example. BL2 (CO1)

[OR]

12. Three machines A,B and C manufacture a product 25%, 55% and 20% respectively. Of their output, 2%,1% and 3% of the products are defective. If a defective product is selected at random from the output, find the probability that it is manufactured by machine
    a) A, b) B, and c) C. [4+3+3] BL4 (CO1)

13. What are Ethical challenges of the data in Data science Problems? BL2(CO2)

[OR]

14. The following are marks of 40 students in a class:
    45,50,55,35,65,52,58,70,63,40,47,52,57,37,67,80,85,54,73,82,
    63,85,37,42,90,67,68,74,76,91,66,43,47,68,72,54,69,76,86,95.
    a) Draw the stem and leaf diagram
    b) Find quartiles and draw Box-Whisker plot. [4+6] BL3 (CO2)

15. How do you select important variables while working on a data set? BL4(CO3)

[OR]

16. a) Give the steps in the KNN machine learning algorithm. BL4 (CO3)
    b)The following is the data on heights (X) and weights(Y) of five students.

X	170	165	174	160	168
Y	70	67	72	62	70

    Fit the regression line of Y on X and predict Y when X=162. BL4 [4+6]

P.T.O

17. Does the recommendation Engine Work in Real time Discuss? BL3(CO4)

[OR]

18. Perform the principal component analysis to the following data. BL6 (CO4)

X_1	5	3	8	6	7
X_2	7	5	5	3	6

19. How do you select the important features while working on a data set? BL5(CO5)

[OR]

20. Discuss the wrapper methods for feature selection? BL6(CO5)

*****