CVR COLLEGE OF ENGINEERING

An UGC Autonomous Institution - Affiliated to JNTUH

B.Tech. **IV** Year **I** Sem. **I Mid Examinations**, October - 2021

Subject**: Fundamentals of Data Science (OE-1)**

Date: 12-10-2021 Time: 1 hr 30 min Max. Marks: **30 M**

Answer **ALL** questions **3x10 = 30 M**

1. Explain the following terms with suitable examples [CO1]
   (a) Data Science (b) Datafication
   (c)Independent and dependent variables (d) Population and sample

(OR)

2. a) In a bolt factory, machines A, B and C manufacture respectively 25%, 30% and 45% of the total production. Of their output 4%, 3% and 1% are known to be defective bolts. [CO1]

   (i) If a bolt is selected at random from the total output, find the probability that it is a defective.
   (ii) If a defective bolt is selected at random from the total output, find the probability that it is produced by machine C.

   b) In a family of 2 children, find the probability that there are [CO1]
   (i) No boys (ii)Exactly one boy(iii) At least one boy (iv) at most one boy

3. What is the purpose of Exploratory Data Analysis in Data Science? Discuss the tools available in EDA? [CO2]

(OR)

4. The following is the frequency table of weights of 100 students. [CO2]

Weight of students (in kg)	20-40	40-60	60-80	80-100	100-120
No. of students	4	27	40	23	6

   (a) Find mean, Standard deviation and coefficient of variation. Also interpret the result.
   (b) Also draw the histogram.

5. a) What is the use of linear regression? Explain the procedure in the method. [CO3]
   b) The following figures relate to advertising expenditure (X) and Sale (Y)[CO3]

X	8	10	12	13	15	16
Y	42	44	47	46	43	45

   (i) Construct regression line of Y on X.
   (ii) Estimate Y when X=18.

(OR)

6. a) Explain the K nearest neighbour algorithm for classification. [CO3
   b) The height in cm (X_1), weight in kg (X_2) and size (Y) of the T-shirt required for 6 customers are given below [CO3]

X1	161	160	163	162	165	170
X2	61	59	61	64	62	65
Y	M	M	M	L	L	L

   A new customer has height 164 cm and weight 63 kg. Predict the size of his T-shirt using KNN method. Here M indicates medium size and L indicates large size.

*****