Total No. of Questions : 4] **SEAT No. :**

PA-10288 **[Total No. of Pages : 1**

[6009]-322

T.E. (Computer Engineering) (Insem.)

DATA SCIENCE AND BIG DATA ANALYTICS

(2019 Pattern) (Semester - II) (310251)

Time : 1 Hour] ***[Max. Marks : 30***

Instructions to the candidates:

1) ***Answer questions Q.1 or Q.2, Q.3 or Q.4.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right side indicate full marks.***
4) ***Assume suitable data if necessary.***
5) ***Use of Scientific calculator is allowed.***

Q1) a) What are dimensionality reduction and its benefits? **[4]**

b) What is data wrangling? Why do you need it? **[5]**

c) What is regression? Explain different types of regression with example. **[6]**

OR

Q2) a) Differentiate between Data Science, Machine Learning and AI. **[4]**

b) What does feature engineering typically includes? **[5]**

c) What is Data Discretization, explain Forms of data discretization. **[6]**

Q3) a) Write a short note on contingency table, explain with example. **[4]**

b) With an example explain Baye's theorem. Also explain its key terms. **[5]**

c) Is there a correlation between the variables in the following data set? **[6]**

Hours	9	15	25	14	10	18	19	16	20	18
Marks	39	56	93	61	50	75	42	70	66	32

OR

Q4) a) What is population & how is it differ from a sample? **[4]**

b) With an example, explain one-tailed & two-tailed t-tests. **[5]**

c) Describe the Chi-Square Test of Independence. **[6]**

◆ ◆ ◆