Total No. of Questions : 4]

SEAT No. : ____

PD103

[Total No. of Pages : 1

[6410] 424

T.E. (Artificial Intelligence and Data Science) (Insem)

DATA SCIENCE

(2019 Pattern) (Semester - II) (317529)

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

Instructions to the candidates:

1) ***Answer Q.1 or Q.2, Q.3 or Q.4.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right indicate full marks.***
4) ***Assume suitable data if necessary.***

Q1) a) Differentiate Structured and Unstructured Data. **[5]**

b) List and explain various Data Discretization techniques. **[5]**

c) What is Data Wrangling? Why data wrangling is important in data science? **[5]**

OR

Q2) a) What is Data Science? How data science is similar to and differ from Information Science? **[5]**

b) Describe Bayes theorem with an example. **[5]**

c) Explain the Data Transformation methods in detail? **[5]**

Q3) a) What are different measures of dispersion? For the given numbers find out variance and standard deviation 4, 34, 11, 12, 2 and 26. **[5]**

b) Here are the 19 scores listed out. 5, 7, 10, 15, 19, 21, 21, 22, 22, 23, 23, 23, 23, 23, 24, 24, 24, 24, 25 Calculate IQR for below the first quartile and above the third quartile. Identify outliers if any? **[5]**

c) Define Type I and Type II Error. Give example to differentiate between the two types of error. **[5]**

OR

Q4) a) What are different measures of central tendency? A pizza outlet overview its weekly sales. They sold 57 cheese pizzas, 63 pasta pizzas, 53 veggies pizzas, 68 cottage cheese pizzas, and 56 max cheese pizzas, Find the mean of all the pizzas sold by them. **[5]**

b) Explain the Chi-Square hypothesis testing with an example. **[5]**

c) What is T-test? What are the types of T-tests? Explain by example the number of variables and degree of freedom. **[5]**