UPES
UNIVERSITY OF TOMORROW

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2022

Course: Applied Data Science   **Semester: 8**
Program: B. Tech. CSE spl. in BigData   **Time   : 03 hrs.**

Course Code: CSDV 4004   **Max. Marks: 100**

Instructions:
1. **All questions are compulsory.**

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Discuss about the different types of data in statistics.	4	CO1
Q 2	Explain confusion matrix and ROC curve.	4	CO5
Q 3	Elaborate Naïve Bayes theorem and its applications.	4	CO2
Q 4	Write the significance of p-value.	4	CO3
Q 5	Discuss the advantages and disadvantages of decision tree algorithm.	4	CO3

SECTION B
(4Qx10M= 40 Marks)

Q 1	Elaborate the different ways to avoid overfitting your model.	10	CO4
Q 2	Explain the random forest algorithm.	10	CO5
Q 3	Discuss Correlation and its following types: Karl Pearson Correlation & Spearmen's rank Order Correlation.	10	CO5

Q 4	Write the steps to test hypothesis.
	OR
	Explain Correlation in Statistics. Discuss the significance of
	  a. Correlation coefficient = 0
	  b. Correlation coefficient = 1
	  c. Correlation coefficient = -1
		10	CO2

SECTION-C
(2Qx20M=40 Marks)

Q 1	Describe the different kinds of independent and dependent variables are needed in conducting a Chi Square test (nominal, ordinal, continuous). Then, create a research question related to Social Work, that you could answer with a Chi Square.	20	CO2

	OR Explain Decision Tree Algorithm and random forest algorithm and draw out the comparison between both.		
Q 2	Explain the term sampling. Discuss the different types of Probability Sampling methods.	**20**	**CO1**