NAVRACHANA UNIVERSITY
a UGC recognized University

School:	School of Engineering and Technology
Program/s:	B. Tech (IT)
Year: 3rd	**Semester:** V
Examination:	End Semester Examination
Examination year:	December - 2021

Course Code: IT319	**Course Name:** Introduction to Data Science		
Date: 9/12/2021		**Total Marks:**	40
Time: 8:30 am to 10:30 am		**Total Pages:**	2

Instructions:

➔ *Write each answer on a new page.*
➔ *Use of a calculator is permitted.*
➔ *Draw all relevant waveforms in answer sheet only.*
➔ ** COs=Course Outcome mapping. # BTL=Bloom's Taxonomy Level mapping*

Q. No.	Details	Marks	COs*	BTL#
Q.1(a)	(1) What is cost function of Logistic Regression Model? (2) Using ___ method, optimal parameters of Multiple Linear Regression Model can be determined. (3) Classification problems can be solved by Logistic regression algorithm. (True / False) (4) Spam filtering is an example of clustering problem. (True / False)	4	CO1	BT1, BT2
Q.1(b)	**Attempt Any Three:** (1) Differentiate between linear regression model and polynomial regression model. (2) Briefly explain clustering methods in unsupervised learning. (3) Write python codes to locate and remove null values in the data. (4) Write formula for F1- score.	6	CO1, CO2, CO3	BT1, BT2, BT5
Q.1 (c)	**Attempt Any TWO:** (1) Consider the following data: <table><tr><td>Y actual</td><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td><td>0</td><td>1</td></tr><tr><td>Y predicted</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr></table> Write True positives, True negatives, False positives, False negatives. Create confusion matrix. Calculate Accuracy, Precision and Recall. (2) Write all the python commands to train (split, train, predict, evaluate) the any data set using Logistic Regression Algorithm. (3) Explain Polynomial Regression Algorithm in Machine Learning.	10	CO1, CO2, CO3	BT1, BT2, BT4, BT5

Q.2(a)	Answer in short:	4		
	(i) Naïve Bayes method can be applied to solve regression problem. (True / False) (ii) Why we need to normalize data before analyze it? (iii) Naïve Bayes classifier is based on ___ theorem of probability theory. (iv) Write python function name to train the model using K- nearest neighbor method in regression problem.		CO2	BT1, BT2
Q.2(b)	Attempt Any THREE:	6		
	(i) Explain how to create confusion matrix in python? (ii) Write piece of python codes to train the model for classification problem using KNN method. (iii) State different distances used in KNN method. (iv) Give two applications of unsupervised learning with proper justification.		CO3, CO4	BT2, BT3, BT5
Q.2 (c)	Attempt ANY TWO: (i) Suppose we take a sample of seven households from a small city and collect information on their incomes and food expenditures for the last month. The information obtained (in hundreds of dollars) is given in following table. Income Food Expenditure 55 14 83 24 38 13 61 16 33 9 49 15 67 17 Write python codes to find the predicted values of food expenditure using Linear Regression Model of ML for the data. (ii) Describe algorithm of Naïve Bayes method to solve classification problem. (iii) Explain python implementation of K-means clustering algorithm.	10	CO1, CO2, CO3	BT1, BT2, BT4, BT5

***********End of Question Paper***********