Title : Question Paper **FF No. 868**

Reg.No. | | | | | | | |

Bansilal Ramnath Agarwal Charitable Trust's
VISHWAKARMA INSTITUTE OF TECHNOLOGY, PUNE – 411037.
(An Autonomous Institute Affiliated to Savitribai Phule Pune University)
Examination: ESE

Year: S.Y. Common **Branch:**

Subject: Data science **Subject Code: MD 2201**

Max. Marks:60 **Total Pages of Question Paper: 1**

Day & Date: Monday, 19/12/22 **Time:** 8.30 am -10.30 am

Instructions to Candidate

1. All questions are compulsory.
2. Neat diagrams must be drawn wherever necessary.
3. Figures to the right indicate full marks.

Q.No.	CO No	BT No		Max marks
Q.1. (A)	1	1	What is Code book or meta data? Explain with an example.	4
(B)	1	1	Explain with examples the terms – raw data and processed data	4
Q. 2. (A)	2	2	Distinguish between point estimate and confidence interval	3
(B)	2	2	What is the importance of significance level? How it regulates the possibility of occurrence of type 1 and type 2 errors?	6
(C)	2	2	How are the Margin of Error and Standard error related with each other?	3
Q. 3. (A)	3	1	State the formula for Lp norm. Show with the help of an example, L1 metric distance is always larger than L2 metric distance	6
(B)	3	2	Draw a typical 'n x n' hessian matrix. How is it used in optimization?	4
Q. 4. (A)	4	4	In a certain regression activity, following scores are obtained. SSR = 18.12, SSE = 2.21. What is the value of R2? Is this regression a good fit?	6
(B)	4	1	What are dichotomous variables in the context of Logistic regression? Give some examples	4
Q. 5. (A)	5	4	Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9). Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2). **OR**	6
Q. 5. (A)	5	4	Apply K-nn and predict the class for the test point (3,7) for k=3. Training points with class are (x,y,class) (7,7,2), (7,4,2), (3,4,1), (1,4,1),(2,5,2),(3,8,1)	
Q.5. (B)	5	3	How do you define Genie impurity and entropy impurity? What will their values be, for the purest node?	4
Q. 6. (A)	6	2	How would you execute the k-fold cross-validation strategy? Why is Leave-one -out method its specialization?	4
(B)	6	3	A Confusion matrix for a classification exercise returns the following values – TP = 0.962, TN = 0.93, FP = 0.12, FN = 0.07. Calculate Accuracy, precision, recall, sensitivity, specificity and f-score	6