U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

January 2024 Semester End Main Examinations

Programme: B.E. **Semester: VII**
Branch: Mechanical Engineering **Duration: 3 hrs.**
Course Code: 23ME7DEMDS **Max Marks: 100**
Course: Mechanical Data Science

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

				CO	PO	Marks
			UNIT - I			
1	a)		Discuss the role of John Tukey in shaping the field of data science and Exploratory Data Analysis.	CO1	PO1	**04**
	b)		Explain the role of Mechanistic Data Science and development of science from data.	CO1	PO1	**06**
	c)		Enumerate the dataset, features and pre-processing for feature-based diamond pricing.	CO1	PO1	**10**
			UNIT - II			
2	a)		Define least squares optimization and its main purpose.	CO2	PO1	**04**
	b)		Explain the method of Least Squares Optimization for Linear Regression.	CO2	PO1	**08**
	c)		What are the analysis and steps involved in Vickers Hardness for Metallic Glasses and Ceramics?	CO2	PO1	**08**
			OR			
3	a)		Define nonlinear regression and explain why it is used when relationships between variables are complex with an example.	CO2	PO1	**04**
	b)		Define optimization and explain the significance of the cost function c(w) in this process.	CO2	PO1	**08**
	c)		Why might linear regression not be suitable for modeling something like bacterial growth, and which type of regression would provide a better fit?	CO2	PO2	**08**
			UNIT - III			
4	a)		Explain the concept of logistic regression, highlight the similarities and differences with linear regression.	CO3	PO1	**08**
	b)		Develop a **decision tree classification** model using python program to categorize the dimensional error an alloy machined using wire EDM from Table Q4b is acceptable or not. Program must include:	CO3	PO2	**12**

			(i) Finding and placing missing values by acceptable method. (ii) Finding best score and parameters: C, kernel, criterion, maximum depth, minimum leaf nodes (use suitable parameters for chosen model/algorithm). (iii) Model development and testing. (iv) Print accuracy and confusion matrix for the best parameter model (v) Can we use bagging or boosting algorithm for this dataset? Justify. **Table Q4b: Dimensional precision of an alloy machined using wire EDM (electrical discharge machining).** Test Run | Pulse-on time (µs) | Pulse-off time (µs) | Current (A) | Bed speed (µm/s) | Dimensional error (µm) | DE acceptable? 1 | 16 | 4 | 3 | 20 | 3 | Yes 2 | 16 | 6 | 4 | 25 | 5 | Yes 3 | 16 | 8 | 5 | 30 | 7 | No 4 | 16 | 10 | 6 | 35 | 8 | No 5 | 20 | NA | 4 | 30 | 4 | Yes 6 | 20 | 6 | 3 | 30 | 5 | Yes 7 | 20 | 8 | 6 | 20 | 8 | No 8 | 20 | 10 | NA | 25 | 9 | No 9 | 24 | 4 | 5 | 35 | 5 | Yes 10 | 24 | 6 | 6 | 30 | 11 | No 11 | 24 | NA | 3 | 25 | 7 | No 12 | 24 | 10 | 4 | 20 | 10 | No 13 | 28 | 4 | 6 | 25 | 8 | No 14 | 28 | 6 | 5 | 20 | 7 | No 15 | 28 | 8 | NA | 35 | 6 | Yes 16 | 28 | 10 | 3 | 30 | 9 | No			
			OR			
5	a)		Illustrate the concept of good decision boundary in SVM using hyperplane.	*CO3*	*PO1*	**08**
	b)		Develop a logistic regression model using python program to identify surface corrosion in copper samples machined using super finish turning is high or low from Table Q5b. Program must calculate the accuracy (both train and test), confusion matrix (both train and test), prediction and probability predictions using the model. Also, write the program to find the optimal threshold. **Table Q5b: Surface characteristics of copper samples undergone super finish turning operation.** Samples | Tool nose radius (mm) | Feed rate (mm/rev) | Lubrication | Surface roughness (µm) | Surface corrosion 1 | 0.8 | 0.05 | With | 2.30 | Low 2 | 0.4 | 0.05 | With | 2.06 | Low 3 | 0.8 | 0.1 | With | 3.76 | High 4 | 0.8 | 0.05 | Without | 2.66 | Low 5 | 0.4 | 0.1 | Without | 6.04 | High 6 | 0.8 | 0.1 | Without | 3.95 | High 7 | 0.4 | 0.05 | Without | 3.46 | High 8 | 0.4 | 0.1 | With | 5.19 | High	*CO3*	*PO2*	**12**
			UNIT - IV			
6	a)		Define clustering in the context of unsupervised learning and explain its goal.	*CO4*	*PO1*	**04**
	b)		Explain the types of linkages used in dissimilarity measures.	*CO4*	*PO1*	**08**
	c)		List the difference between agglomerative and divisive clusterings.	*CO4*	*PO1*	**08**
			UNIT - V			
7	a)		Mention the steps in Artificial Neuron Network.	*CO5*	*PO1*	**04**

		b)	Explain the function of biological neuron with the help of schematic diagram.	*CO5*	*PO1*	**06**
		c)	Define CNN and explain the fundamental architecture of CNN with the help of schematic diagram.	*CO5*	*PO1*	**10**

****