U.S.N.										

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

Autonomous Institute Affiliated to VTU

June 2025 Semester End Main Examinations

Programme: B.E. **Semester: VI/V**
Branch: Information Science and Engineering **Duration: 3 hrs.**
Course Code: 22IS6PCMLG **Max Marks: 100**
Course: Machine Learning

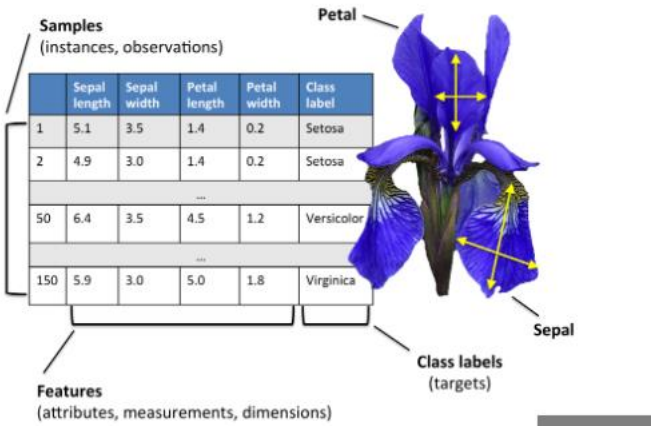
Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

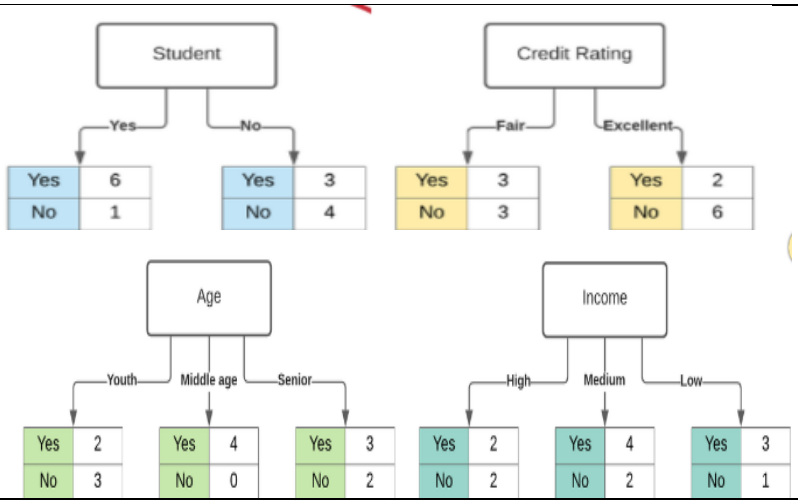
Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT - I	CO	PO	Marks
1	a)	With suitable diagrams distinguish how a spam filter model built using machine learning technique is better compared to a traditional programming technique	CO1		10
	b)	Consider the Employee Data set and design the steps involved to transform the raw data into an understandable and readable format using the concept of data pre-processing using Sklearn libraries.	CO2	PO1	10
		OR			
2	a)	If your model performs great on the training data but generalizes poorly to new instances, what is going wrong? Elaborate on the methods to resolve this.	CO1		4
	b)	Discuss GridSearch and Randomized Search along with suitable python code using Sklearn libraries.	CO2	PO1	8
	c)	Main task of machine learning is to select a learning algorithm and train it on some data, the two things that can go wrong are "bad algorithm" and "bad data." Explain the main challenges with respect to "bad algorithm" and "bad data."	CO1		8
		UNIT - II			
3	a)	Consider a car dataset that classifies the vehicle into four types: Sedan, HatchBack, Sports car and MiniVan. Explain with suitable	CO1		6

Row no	State	Salary	Yrs of Experience
1	NY	57400	Mid
2	TX		Entry
3	NJ	90000	High
4	VT	36900	Entry
5	TX		Mid
6	CA	76600	High
7	NY	85000	High
8	CA		Entry
9	CT	45000	Entry

Missing values

			diagrams as to how OVA and OVO techniques can solve classification problem			
		b)	Illustrate how ROC curve shows the trade-off between sensitivity (TPR) and specificity (1 – FPR) based on AUC	*CO2*	*PO1*	**4**
		c)	Design the following steps involved in building the Logistics Regression model using Sklearn libraries breast cancer dataset, • Import Libraries and the dataset • Splitting the dataset into the Training set and Test set with test size =20% • Feature Scaling -Apply Standard Scalar on complete X-train and X-test • Training the Logistic Regression model on the Training set . Print the confusion matrix, accuracy score and classification report.	*CO2*	*PO1*	**10**
			OR			
	4	a)	Design a Polynomial Regression model using Sklearn libraries considering Insurance.csv to calculate Insurance Premium based on age. Split the dataset with test size =30% and apply Standard Scalar as feature Scaling technique and print MAE,RMSE the performance measure.	*CO3*	*PO2*	**10**
		b)	The dataset has four features – sepal length, sepal width, petal length and petal width with 150 rows that classify each instance as either Setosa, Versicolor, Virginica as shown in below diagram. Design a machine learning model using Sklearn libraries that identifies it as Setosa or not along with necessary pre-processing steps and performance measures.  	*CO3*	*PO2*	**10**
			UNIT - III			
	5	a)	With suitable examples, explain the different regularization hyper parameters used for restricting the shape of the decision tree.	*CO1*		**5**
		b)	Explain CART algorithm with cost function used for classification	*CO1*		**5**
		c)	Consider you want to build a decision tree classifier that determines if a person will purchase a credit card or not . The classifier includes features such as age, income, credit rating, and a student. Find the best feature for the first and second split of the tree – the root node and depth 1 leaf nodes in the decision tree.	*CO2*	*PO1*	**10**



			OR			
6	a)	Design a Decision Tree model using SKlearn library for the petrol consumption dataset that has 4 features- petrol_tax, average_income,paved_highways and population_delivery which predicts petrolconsumption. Apply necessary pre-processing steps and performance measures. Predict the petrol consumptionfor values - [9,3471,1250,0.58].	***CO3***	***PO2***	**10**
	b)	What role does the Gini index play in minimizing misclassification errors in decision tree models?Calculate Gini Index for **past trend, open interest and trading volume .**	***CO2***	***PO1***	**10**

Past Trend	Open Interest	Trading Volume	Return
Positive	Low	High	Up
Negative	High	Low	Down
Positive	Low	High	Up
Positive	High	High	Up
Negative	Low	High	Down
Positive	Low	Low	Down
Negative	High	High	Down
Negative	Low	High	Down
Positive	Low	Low	Down
Positive	High	High	Up

		UNIT - IV			
7	a)	Demonstrate with an example how Ada boosting supports to train predictors sequentially.	***CO1***		**10**
	b)	Design a machine learning model using Sklearn libraries for MNIST dataset by considering first 60,000 instances for training, and remaining 10,000 instances for testing. Apply dimensionality reduction to obtain principal components with variance ratio of 95 and use suitable model for classification.	***CO3***	***PO2***	**10**
		OR			
8	a)	Illustrate with an example the iterative process of refining predictors in Gradient Boosting and how it leads to better predictive performance. How does Gradient Boosting differ from AdaBoost in terms of handling prediction errors?	***CO1***		**10**
	b)	Discuss the need of Ensemble models. List and explain the different types Voting methods.	***CO1***	***PO1***	**5**

		c)	What are the two main approaches for dimensionality reduction? Explain in detail with necessary diagrams.	*CO1*	*PO1*	**5**
			UNIT - V			
	9	a)	Elucidate unsupervised learning? How unsupervised learning is different from supervised learning. Mention any 3 applications where you can apply unsupervised learning model.	*CO1*		**10**
		b)	Considering below dataset, design the model involved in clustering the gender based on the similarity using necessary SKlearn libraries. **Customer ID** | **Gender** | **Age** | **Annual Income (k$)** | **Spending Score (1-100)** 1 | Male | 19 | 15 | 39 2 | Male | 21 | 15 | 81 3 | Female | 20 | 16 | 6 4 | Female | 23 | 16 | 77 5 | Female | 31 | 17 | 40 6 | Female | 22 | 17 | 76	*CO3*	*PO2*	**10**
			OR			
	10	a)	Given a dataset, a radius value (epsilon), and a minimum number of points (MinPts), how are core, border, and noise points defined by DBSCAN. Explain with an example. In what way does DBSCAN use those points to cluster the dataset?	*CO1*		**10**
		b)	Apply K (=2)-Means algorithm over the data (185, 72), (170, 56), (168, 60), (179,68), (182,72), (188,77) up to two iterations and show the clusters. Initially choose first two objects as initial centroids.	*CO3*	*PO2*	**10**

******