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: IV**
Branch: CSE(DS), AI&DS **Duration: 3 hrs.**
Course Code: 23DS4PCMLG **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.

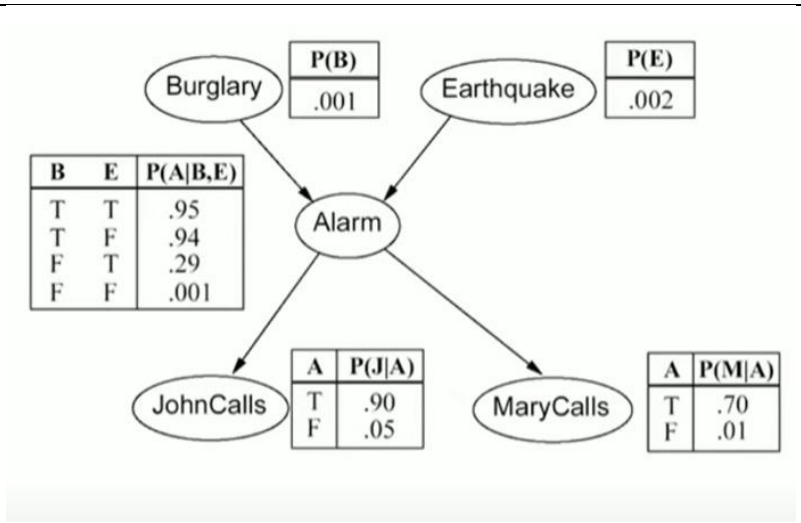
		UNIT - I	CO	PO	Marks
1	a)	Elucidate on types of Machine Learning with appropriate examples.	CO1	PO1	5
	b)	Build a decision tree using ID3 to predict Buys computer (Yes/No) based on the other attributes for the data given in Table 1.	CO2	PO2	10
	c)	Highlight challenges in decision tree learning and provide mitigation strategy for any two.	CO2	PO2	5
		OR			
2	a)	Differentiate between training, validation, and testing data in Machine Learning.	CO2	PO2	5
	b)	Determine the student with the attribute values (Bachelors, Medium, Python) will be hired or not based on the ID3 algorithm that has trained the model using the data set provided in Table 2	CO3	PO3	10

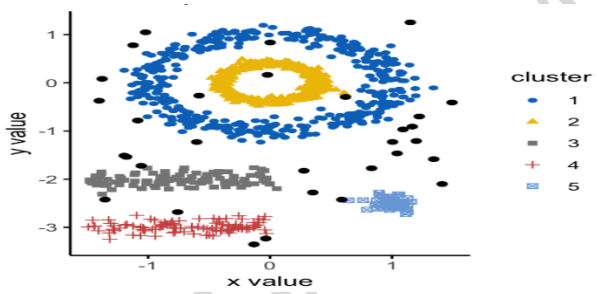
Table 1

ID	Age	Income	Student	Credit_rating	Buys_computer
1	<=30	High	No	Fair	No
2	<=30	High	No	Excellent	No
3	31–40	High	No	Fair	Yes
4	>40	Medium	No	Fair	Yes
5	>40	Low	Yes	Fair	Yes
6	>40	Low	Yes	Excellent	No

ID	Education	Experience	Skills	Hired
1	Bachelors	Low	Python	No
2	Bachelors	High	Python	Yes
3	Masters	Medium	Java	Yes
4	Masters	Low	Python	No
5	PhD	High	Java	Yes
6	Bachelors	Medium	Java	No

		c)	Outline the core mechanism of the CART algorithm for decision tree construction.	*CO2*	*PO2*	**5**
			UNIT - II			
	3	a)	Discuss how a Non-Linear SVM handles complex data.	*CO2*	*PO2*	**6**
		b)	A dataset contains points in a 3D space (x, y, z) with their respective classes: • Point A: (1, 2, 1), Class: Red • Point B: (2, 2, 2), Class: Blue • Point C: (3, 1, 3), Class: Red • Point D: (4, 3, 2), Class: Blue • Point E: (2, 4, 3), Class: Red Determine the class for a new point P: (2, 3, 2) which needs to be classified using k-NN with k = 4.	*CO3*	*PO3*	**10**
		c)	Differentiate between hard margin and soft margin SVM.	*CO2*	*PO2*	**4**
			OR			
	4	a)	Discuss the challenges that arise when using SVMs on imbalanced datasets.	*CO2*	*PO2*	**6**
		b)	A dataset contains points with features (Age, Income in thousands, Hours Worked per Week) and a class label (Promoted: Yes/No): • Employee 1: (25, 30, 40), Class: No • Employee 2: (30, 50, 35), Class: Yes • Employee 3: (35, 40, 45), Class: No • Employee 4: (28, 60, 50), Class: Yes • Employee 5: (40, 45, 30), Class: Yes A new employee P: (27, 45, 42) is to be classified using k-NN with k = 3, but the features have different scales. First, normalize the features using min-max normalization (to scale them to [0, 1]), then proceed.	*CO3*	*PO3*	**10**
		c)	Discuss the key limitations of the k-NN algorithm.	*CO2*	*PO2*	**4**
			UNIT - III			
	5	a)	Differentiate between Maximum Likelihood and Minimum Description Length (MDL) principles.	*CO2*	*PO2*	**6**

		b)	 What is the probability that the alarm has sounded but neither a burglary nor an earthquake has occurred, both John and Mary call? Write Python Program for the same.	*CO2*	*PO2*	**6**
		c)	Outline the advantages, disadvantages, and applications of Expectation Maximization Algorithm.	*CO2*	*PO2*	**8**
			OR			
	6	a)	Describe the Bayes Optimal Classifier and contrast it with the Naïve Bayes Classifier.	*CO3*	*PO3*	**6**
		b)	Discuss how Bayesian Learning handles noise and uncertainty better than other methods?	*CO3*	*PO3*	**8**
		c)	A diagnostic test for a disease has the following statistics: • P(Disease)=0.01 P(No Disease)=0.99 • P (Positive Test | Disease) =0.95 • P (Positive Test | No Disease) =0.1 Determine the probability that a person has the disease given a positive test result**?**	*CO4*	*PO4*	**6**
			UNIT - IV			
	7	a)	With the help of python program describe the following ensemble methods in machine learning: i) Bagging and Pasting ii) Voting Classifiers	*CO3*	*PO3*	**10**
		b)	Describe the architecture and working of a Random Forest. Discuss how it differs from a single Decision Tree.	*CO2*	*PO2*	**10**
			OR			
	8	a)	Discuss and analyze the following boosting algorithms with suitable explanations and illustrative code examples:	*CO2*	*PO2*	**10**

			i)AdaBoost ii) Gradient Boosting			
		b)	Describe the concept of Stacking in ensemble learning. How does it differ from Bagging and Boosting in terms of model training and prediction? Provide a practical use case.	*CO3*	*PO3*	**10**
			UNIT - V			
	9	a)	I. Write the python code for the following i. DBSCAN for the given data instances shown in the figure ii. To show label of the data instances iii. To obtain length of the core sample indices iv. To obtain components of the data instances. v. Assume eps = 0.05 for the figure 1, analyze DBSCAN when eps = 0.20. II. Enumerate DBSCAN algorithm 	*CO3*	*PO3*	**10**
		b)	Describe how clustering can be useful in semi-supervised learning. What role does it play when labels are partially available?	*CO3*	*PO3*	**5**
		c)	Discuss the learning task and Q learning in the context of reinforcement learning.	*CO2*	*PO2*	**5**
			OR			
	10	a)	A robot can move in a grid where each cell represents a state. It can move left, right, up, or down. The robot gets a reward of +1 at the goal and 0 otherwise. Model this as a Markov Decision Process. Specify the states, actions, rewards, and transitions.	*CO4*	*PO4*	**8**
		b)	Discuss how DBSCAN identifies core, border, and noise points? Briefly explain with an example.	*CO2*	*PO2*	**6**
		c)	How can clustering algorithms be used for anomaly detection? Illustrate with a suitable example or scenario.	*CO2*	*PO2*	**6**

******