U.S.N.										

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

Autonomous Institute Affiliated to VTU

January / February 2025 Semester End Main Examinations

Programme: B.E. **Semester: V**

Branch: CSE (IoT & Cybersecurity including Blockchain) **Duration: 3 hrs.**

Course Code: 23IC5PCMLG **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)		Identify the type of machine learning technique for the following problem statements i. Photo-hosting service such as Google Photos. ii. A gallery has images labelled as cats and dogs. iii. An action game rewards his action hero with + 10 as a reward -15 as a penalty. iv. Similarity between two emails is calculated and marked as spam. v. SVM model is applied to train the patient dataset.	CO2	PO 2	**05**
	b)		Construct a decision tree using the ID3 algorithm for the following data.	CO3	PO3	**08**

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

	c)		Imagine you are working on a classification task for a financial institution to predict whether a loan applicant is likely to default. You decide to use a decision tree as your predictive model. After training the tree, you notice that it perfectly classifies the training data, but performs poorly on the validation set, indicating overfitting. Identify and enumerate on approaches to avoiding overfitting in decision tree learning and criterion used to determine the correct final tree size.	CO2	PO2	**07**
			OR			

	2	a)	Imagine you are the head of a cybersecurity team for an email service provider. Your team has been receiving numerous complaints from users about an increasing number of spam emails that are making it into their inboxes. The traditional rule-based filters your company has been using seem to be less effective against these sophisticated spam campaigns. Given this scenario, how would implementing machine learning improve the accuracy and efficiency of spam detection compared to the current rule-based system?	*CO2*	*PO2*	**06**
		b)	Consider the training data with major, experience, percentage, communication as the attributes. Hired is the target variable with two classes: Yes/No	*CO3*	*PO3*	**10**

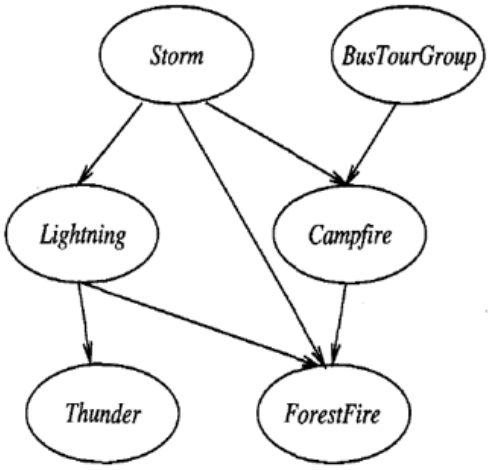
Major	Experience	Percentage	Communication	Hired?
CS	Programming	>75 <85	Good	Yes
CS	Programming	>75 <85	Good	Yes
CS	Management	>75 <85	Good	No
CS	Management	>65<=75	Average	No
Business	Programming	>65<=75	Average	No
Business	Programming	<65	Average	No
Business	Management	>75 <85	Excellent	Yes
CS	Management	>75 <85	Excellent	Yes
CS	Management	>75 <85	Excellent	Yes
Business	Programming	<65	Average	No
CS	Programming	>75 <85	Excellent	No

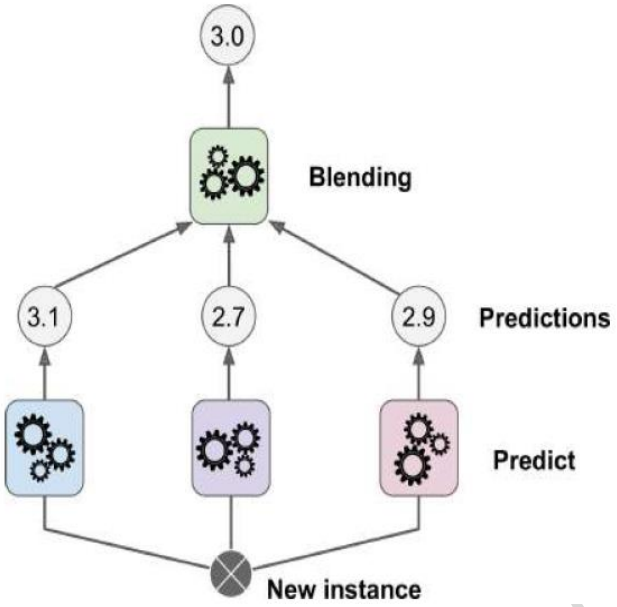
i) Calculate the Information Gain for the entire data set

ii) Calculate the Gini for Percentage and Communication of the candidate.

		c)	Describe how the CART algorithm determines the best split at each node. What criteria does it use for classification tasks?	*CO1*	*PO1*	**04**
			UNIT - II			
	3	a)	Enumerate on Linear Support Vector Machines and distinguish between soft margin and hard margin.	*CO2*	*PO2*	**06**
		b)	Given the dataset, predict the outcome of an instance. The features are BMI and Age, and the target variable is Diabetic. NOTE: Diabetic = 1 indicates the patient is diabetic, and 0 indicates non-diabetic. Solve using the k-NN algorithm. Assume k=3.	*CO3*	*PO3*	**10**

			<table><tr><th>Sl. No.</th><th>BMI</th><th>Age</th><th>Diabetic</th></tr><tr><td>1</td><td>33.6</td><td>50</td><td>1</td></tr><tr><td>2</td><td>26.6</td><td>30</td><td>0</td></tr><tr><td>3</td><td>23.4</td><td>40</td><td>0</td></tr><tr><td>4</td><td>43.1</td><td>67</td><td>0</td></tr><tr><td>5</td><td>35.3</td><td>23</td><td>1</td></tr><tr><td>6</td><td>35.9</td><td>67</td><td>1</td></tr><tr><td>7</td><td>36.7</td><td>45</td><td>1</td></tr><tr><td>8</td><td>25.7</td><td>46</td><td>0</td></tr><tr><td>9</td><td>23.3</td><td>29</td><td>0</td></tr><tr><td>10</td><td>31</td><td>56</td><td>1</td></tr><tr><td>11</td><td>43.6</td><td>40</td><td>?</td></tr></table>			
	c)	Elucidate on Instance-based Learning.	*CO2*	*PO2*	**04**
		OR			
4	a)	With respect to Nonlinear SVM Classification, explain Polynomial Kernel and Gaussian RBF Kernel with example.	*CO2*	*PO2*	**06**
	b)	<table><tr><th>Height (CM)</th><th>Weight (KG)</th><th>Class</th></tr><tr><td>167</td><td>51</td><td>Underweight</td></tr><tr><td>182</td><td>62</td><td>Normal</td></tr><tr><td>176</td><td>69</td><td>Normal</td></tr><tr><td>173</td><td>64</td><td>Normal</td></tr><tr><td>172</td><td>65</td><td>Normal</td></tr><tr><td>174</td><td>56</td><td>Underweight</td></tr><tr><td>169</td><td>58</td><td>Normal</td></tr><tr><td>173</td><td>57</td><td>Normal</td></tr><tr><td>170</td><td>55</td><td>Normal</td></tr><tr><td>170</td><td>57</td><td>?</td></tr></table> Write the algorithm for KNN and calculate KNN for the above dataset. Consider the K = 5.	*CO3*	*PO3*	**10**
	c)	You are designing a machine learning model to predict house prices based on features like location, size, number of rooms, and age of the property. You decide to use a regression model for this task. What should the training objective of your regression model be? How would you formulate the loss function to minimize prediction errors?	*CO1*	*PO1*	**04**
		UNIT - III			
5	a)	Summarize on Bayesian Learning with the following phenomena. i. Relevance ii. Features iii. Use cases iv. Challenges	*CO1*	*PO1*	**08**

		b)	Given the dataset, predict the outcome of an instance. NOTE: The target variable is Species. Solve using the Naïve Bayes Classifier. **Sl. No. | Color | Legs | Height | Smelly | Species** 1 | White | 3 | Short | Yes | M 2 | Green | 2 | Tall | No | M 3 | Green | 3 | Short | Yes | M 4 | White | 3 | Short | Yes | M 5 | Green | 2 | Short | No | H 6 | White | 2 | Tall | No | H 7 | White | 2 | Tall | No | H 8 | White | 2 | Short | Yes | H 9 | Green | 2 | Tall | No | ?	*CO3*	*PO3*	**08**
		c)	Outline the advantages, disadvantages, and applications of Expectation Maximization Algorithm.	*CO1*	*PO1*	**04**
			OR			
	6	a)	Enumerate on i. Bayes Optimal Classifier ii. GIBBS Algorithm Illustrate with an example on when to choose (i) and (ii).	*CO1*	*PO1*	**10**
		b)	Analyze the Bayesian Belief Network given below. Consider the node: Campfire and expound on conditional dependence/ independence.  	*CO2*	*PO2*	**10**
			UNIT - IV			
	7	a)	i. Illustrate Hard voting and soft voting classifiers with example. ii. Write a pseudocode (use Python libraries) to demonstrate the working of Voting classifiers.	*CO1*	*PO1*	**10**

		b)	Examine in detail the idea of Stacking in Machine Learning using the flow diagram below.  	*CO2*	*PO2*	**10**
			OR			
	8	a)	Identify the difference between bagging and boosting.	*CO1*	*PO1*	**10**
		b)	Demonstrate the working of Random Forest algorithm with a suitable example.	*CO2*	*PO2*	**10**
			UNIT - V			
	9	a)	Enumerate how the Reinforcement learning problem is different from other function approximation tasks.	*CO1*	*PO1*	**08**
		b)	Explain the working of DBSCAN algorithm with suitable code snippet.	*CO1*	*PO1*	**08**
		c)	Construct the Q learning algorithm.	*CO2*	*PO2*	**04**
			OR			
	10	a)	i) Explain the steps involved in k-means algorithm. ii) Identify and explain the technique used to choose k value.	*CO1*	*PO1*	**08**
		b)	Enumerate any two applications of clustering.	*CO1*	*PO1*	**04**
		c)	For the given data, compute two clusters using K-means algorithm for clustering where initial cluster centers are (1.0, 1.0) and (5.0, 7.0). Execute for two iterations. Record Number | A | B R1 | 1.0 | 1.0 R2 | 1.5 | 2.0 R3 | 3.0 | 4.0 R4 | 5.0 | 7.0 R5 | 3.5 | 5.0 R6 | 4.5 | 5.0 R7 | 3.5 | 4.5	*CO2*	*PO2*	**08**

******