Reg.No.:												

VIVEKANANDHA COLLEGE OF ENGINEERING FOR WOMEN
[AUTONOMOUS INSTITUTION AFFILIATED TO ANNA UNIVERSITY, CHENNAI]
Elayampalayam – 637 205, Tiruchengode, Namakkal Dt., Tamil Nadu.

Question Paper Code: 60035

M.E. / M.Tech. DEGREE END-SEMESTER EXAMINATIONS – FEB. 2025
First Semester
Information Technology
P23IT102 – MACHINE LEARNING ALGORITHMS
(Regulation 2023)

Time: Three Hours　　　　Maximum: 100 Marks

Answer ALL the questions

Knowledge Levels (KL)	K1 – Remembering	K3 – Applying	K5 - Evaluating
	K2 – Understanding	K4 – Analyzing	K6 - Creating

PART – A

(10 x 2 = 20 Marks)

Q.No.	Questions	Marks	KL	CO
1.	What is supervised learning, and how does it differ from unsupervised learning?	2	K2	CO1
2.	How does self-training work in semi-supervised learning, and when would you use it?	2	K3	CO1
3.	What is dimensionality reduction, and why is it important?	2	K2	CO2
4.	What are the different ways of data scaling and normalization?	2	K1	CO2
5.	How do Markov Random Fields use the concept of conditional independence between variables in the network?	2	K3	CO3
6.	In which applications do we use linear regression and in which we use logistic regression?	2	K3	CO3
7.	How do Decision Trees handle imbalanced datasets where one class significantly outweighs the other?	2	K3	CO4
8.	How do you choose the optimal number of clusters K in K-Means clustering?	2	K2	CO4
9.	Explain how recommendation Systems are used in e-commerce platform.	2	K3	CO5
10.	What is temporal topic modeling, and why is it important?	2	K2	CO5

PART – B

(5 x 13 = 65 Marks)

Q.No.		Questions	Marks	KL	CO
11.	a)	Explain how deep learning models can be inspired by biological systems to create adaptive systems. Discuss the key biological mechanisms that influence the architecture of modern deep learning systems, and describe specific examples where bio-inspired approaches enhance model adaptability.	13	K3	CO1
		(OR)			
	b)	Discuss the key principles behind statistical learning approaches in machine learning. Explain the mathematical foundations that underpin these models, the role of probability theory, and how statistical learning differs from other learning paradigms. Provide examples of algorithms, their applications, and the challenges associated with statistical learning.	13	K3	CO1
12.	a)	Discuss best practices for splitting data, handling imbalanced datasets, and evaluating model performance. Provide specific techniques for preparing data for both training and testing phases, and explain how to avoid common pitfalls like data leakage and overfitting.	13	K3	CO2
		(OR)			
	b)	You are given the following dataset of 5 samples with 2 features: $X=\begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 2 \\ 2 & 2 \\ 3 & 3 \end{bmatrix}$ Compute the mean and standard deviation for each feature, and use this information to standardize the dataset. Calculate the covariance matrix for the standardized data. Compute the eigenvalues and eigenvectors of the covariance matrix.	13	K3	CO2
13.	a)	What are cliques in the context of Markov Random Fields, and how do clique potentials define interactions within the network? Also, explain how the joint probability distribution of an MRF can be factorized using clique potentials.	13	K3	CO3
		(OR)			

		Marks	KL	CO
b)	Describe the process of linear regression and its application using scikit-learn. How can polynomial regression extend the linear model?	13	K3	CO3
14. a)	You apply K-Means clustering and hierarchical clustering to the same dataset. The K-Means algorithm gives you 3 clusters, while hierarchical clustering suggests 4 clusters. Explain the possible reasons for the difference in the number of clusters and how would you decide which clustering result is better?	13	K3	CO4
	(OR)			
b)	Compare the CART (Classification and Regression Trees), and ID3 algorithms in terms of their splitting criteria, handling of continuous features, and performance on large datasets.	13	K3	CO4
15. a)	Explain the concepts of deep learning and how TensorFlow is utilized in developing deep learning models.	13	K2	CO5
	(OR)			
b)	You have built two topic models: one using LDA and another using NMF. The LDA model yields 5 topics while the NMF model yields 4 topics. Discuss how you would compare the effectiveness of both models in terms of interpretability and coherence. What factors would you consider?	13	K3	CO5

PART – C

(1 x 15 = 15 Marks)

Q.No.	Questions	Marks	KL	CO
16. a)	Imagine you are a data scientist working for a financial institution tasked with predicting loan defaults based on historical customer data. The dataset contains 100,000 records with features such as customer demographics, loan details, payment history, and account information. Your goal is to create a robust predictive model that accurately identifies customers at risk of defaulting on their loans. I. Identify at least five different machine learning algorithms suitable for the loan default prediction task. II. Any critical hyperparameters you would tune during the modelling process and how they could impact the model's performance.	15	K5	CO4

III. Strategies for handling missing values, including techniques such as imputation or deletion, and the rationale for your choice.

IV. Techniques for creating new features from existing data (e.g., interaction terms, aggregations over time) to improve model performance.

V. How you would split the dataset into training, validation, and test sets, including the rationale behind your chosen ratios (e.g., 70/15/15).

(OR)

b) Explain how ensemble learning techniques like Random Forest can improve classification performance, providing examples of their real-world applications.	15	K3	CO3