Reg. No.															

Question Paper Code	14041

B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2025

Third Semester

Artificial Intelligence and Data Science

24AIPC301 - INTRODUCTION TO MACHINE LEARNING

Regulations - 2024

Duration: 3 Hours **Max. Marks: 100**

PART - A (MCQ) (10 × 1 = 10 Marks)
Answer ALL Questions

		Marks	K – Level	CO
1.	Predict which type of learning is most suitable for credit card fraud detection. (a) Unsupervised learning (b) Reinforcement learning (c) Supervised learning (d) Semi-supervised learning	1	K1	CO1
2.	Infer why unsupervised learning is useful in insurance analytics. (a) It predicts claim amounts (b) It finds patterns among policyholders without labels (c) It rewards successful claims (d) It trains models using historical claim labels	1	K2	CO1
3.	If a dataset contains gender (Male/Female) and salary (in USD), which types of data are included? (a) Only categorical data (b) Only numerical data (c) Both categorical and numerical data (d) Neither categorical nor numerical data	1	K2	CO2
4.	Find the purpose of model evaluation metrics in machine learning. (a) To compare different models and assess predictive performance (b) To train the model efficiently (c) To reduce data size (d) To collect more data	1	K1	CO2
5.	Which step of feature engineering focuses on deriving new variables from existing ones? (a) Feature extraction (b) Feature reduction (c) Feature selection (d) Feature construction	1	K1	CO3
6.	How does SVM classify data compared to Decision Trees? (a) SVM uses probability estimation, Decision Trees use distance metrics (b) SVM uses hyperplanes to separate classes, while Decision Trees split data based on features (c) Both use identical mathematical models (d) SVM relies on tree nodes for prediction	1	K2	CO3
7.	In multiple linear regression, the model predicts the dependent variable using: (a) A single independent variable (b) Multiple independent variables (c) Only categorical variables (d) Only numerical variables	1	K1	CO4
8.	Identify the regularization method that can set some coefficients exactly to zero. (a) Ridge Regression (b) Elastic Net Regression (c) Linear Regression (d) Lasso Regression	1	K2	CO4
9.	In a K-Medoids clustering output, if one medoid is very different from its assigned points, what can be inferred? (a) The number of clusters is too small (b) The similarity measure is inappropriate (c) The data is linearly separable (d) The medoid was correctly chosen	1	K2	CO5
10.	In association rule learning, support and confidence are used to (a) Evaluate rule significance (b) Clean the dataset (c) Split the data into clusters (d) Reduce overfitting	1	K1	CO6

PART - B (12 × 2 = 24 Marks)
Answer ALL Questions

11.	Interpret how supervised learning models learn from labeled datasets.	*2*	*K2*	*CO1*
12.	State any four applications of Machine Learning.	*2*	*K1*	*CO1*
13.	Mention the essential best practices to be considered in data preprocessing.	*2*	*K1*	*CO2*
14.	Identify the major performance metrics used to evaluate feature models.	*2*	*K2*	*CO2*
15.	List out the benefits and limitations of feature engineering.	*2*	*K1*	*CO3*
16.	Infer how feature scaling influences the performance of the KNN classifier.	*2*	*K2*	*CO3*
17.	Compare and contrast linear regression and logistic regression.	*2*	*K2*	*CO4*
18.	Explain the principle of univariate and multivariate linear regression techniques.	*2*	*K2*	*CO4*
19.	What is meant by a partitioning clustering method?	*2*	*K1*	*CO5*
20.	Interpret how density-based clustering methods form clusters.	*2*	*K2*	*CO5*
21.	Define representation learning and its purpose.	*2*	*K1*	*CO6*
22.	Summarize how autoencoders contribute to unsupervised feature learning.	*2*	*K2*	*CO6*

PART - C (6 × 11 = 66 Marks)
Answer ALL Questions

23.	a)	Explain the concept of reinforcement learning and discuss the key characteristics and challenges associated with reinforcement-based decision-making problems.	*11*	*K2*	*CO1*
		OR			
	b)	Explain how machine learning techniques contribute to the improvement of healthcare applications, and outline the key benefits and challenges of using them in healthcare solutions.	*11*	*K2*	*CO1*
24.	a)	Illustrate the different types of data used in machine learning and explain how to explore the structure of a dataset with an example.	*11*	*K2*	*CO2*
		OR			
	b)	Discuss in detail the steps for model selection and how does that reflect in the training model of determining an efficient solution for an application of specific domain.	*11*	*K2*	*CO2*
25.	a)	Organize the steps of feature engineering phases and illustrate with example the feature transformation and feature subset selection process.	*11*	*K3*	*CO3*
		OR			
	b)	Outline the ID3 algorithm used for inducing decision tree from the training tuples. Also list down the different attribute selection measures used in the process of decision tree constructions.	*11*	*K3*	*CO3*

Day	Outlook	Humidity	Wind
D1	Sunny	High	Weak
D2	Sunny	High	Strong
D3	Overcast	High	Weak
D4	Rain	High	Weak
D5	Rain	Normal	Weak
D6	Rain	Normal	Strong
D7	Overcast	Normal	Strong

D8	Sunny	High	Weak
D9	Sunny	Normal	Weak
D10	Rain	Normal	Weak
D11	Sunny	Normal	Strong
D12	Overcast	High	Strong
D13	Overcast	Normal	Weak

26. a) Interpret the step-by-step process involved in building a simple linear regression model and examine how the regression line is fitted to the data using the least squares method. *11 K2 CO4*

OR

b) Compare and contrast Ridge, Lasso, and Elastic Net regression techniques, and analyze how each method controls model complexity and prevents overfitting. *11 K2 CO4*

27. a) Examine the suitability of different clustering approaches (partitioning, hierarchical, and density-based) for datasets with varying shapes, sizes, and noise levels. *11 K5 CO5*

OR

b) Assess the advantages and drawbacks of the Apriori algorithm for association rule learning and suggest improvements to enhance its efficiency. *11 K5 CO5*

28. a) Explain the various types of learning beyond supervised and unsupervised learning with suitable examples. *11 K2 CO6*

OR

b) Illustrate the concept of ensemble learning and explain how bagging and boosting works with their key advantages and disadvantages. *11 K2 CO6*