CBCS SCHEME

USN | | | | | | | | | | |

MMC201

Second Semester MCA Degree Examination, Dec.2025/Jan.2026
Machine Learning and Data Analytics Using Python

Time: 3 hrs. Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.
2. M : Marks , L: Bloom's level , C: Course outcomes.

		Module – 1	M	L	C
Q.1	a.	Define supervised, unsupervised, and reinforcement learning with suitable real-life examples.	10	L2	CO1
	b.	List any four Python libraries used for data analysis. Explain their roles with examples.	06	L2	CO1
	c.	Develop a Python program using Pandas and Matplotlib to visualize a dataset (bar/line plot).	04	L3	CO1
		OR			
Q.2	a.	Define Machine Learning. Explain any three real-world applications of Machine Learning.	05	L2	CO1
	b.	Describe the role of Python in data science and machine learning.	05	L2	CO1
	c.	Demonstrate the use of Pandas for data manipulation and visualization using a sample dataset.	10	L3	CO1
		Module – 2			
Q.3	a.	Compare Logistic Regression, KNN, and Decision Trees in terms of working and applications.	10	L4	CO2
	b.	Explain standardization and normalization with examples. Explain the use cases for these approaches.	06	L2	CO2
	c.	Develop a Python code to demonstrate handling missing values in a dataset using Pandas.	04	L3	CO2
		OR			
Q.4	a.	What is feature selection? Explain any two filter-based techniques used for selecting features.	05	L2	CO2
	b.	Develop a Python program to implement Linear Regression using scikit-learn.	05	L3	CO2
	c.	Elucidate the concept of train-test split, cross-validation, and hyperparameter tuning using Grid Search CV.	10	L4	CO2
		Module – 3			
Q.5	a.	Describe Linear and Polynomial Regression. Differentiate between these 2 regression methods.	05	L2	CO3
	b.	Develop a Python program to implement Linear Regression using scikit-learn.	05	L3	CO3
	c.	Explain the significance of evaluation metrics: MAE, MSE, RMSE in regression analysis.	05	L2	CO3
	d.	Develop Python code to build and evaluate a Polynomial Regression model using scikit-learn.	05	L3	CO3

		OR			
Q.6	a.	Explain the working of the K-Nearest Neighbors (KNN) algorithm for classification with an example.	05	L2	CO3
	b.	Develop a Python program to apply Random Forest classification on any dataset and evaluate its performance.	05	L3	CO3
	c.	Explain the working of Decision Tree and Random Forest algorithms with diagrams.	05	L2	CO3
	d.	Compare evaluation metrics: Accuracy, Precision, Recall, F1-score, and ROC-AUC with examples.	05	L3	CO3
		Module – 4			
Q.7	a.	Explain Train-Test split and k-fold cross-validation techniques for model evaluation.	05	L2	CO4
	b.	Define SVM. Differentiate between linear vs. non-linear SVM	05	L2	CO4
	c.	Explain the working of ensemble techniques: Bagging, Boosting, and XGBoost. Use examples to highlight differences.	10	L2	CO4
		OR			
Q.8	a.	What is over fitting and under fitting? How can they be detected and prevented?	05	L2	CO4
	b.	Describe the concept of hyperparameter tuning using Grid Search CV with an example.	05	L2	CO4
	c.	Describe the architecture of Convolutional Neural Networks (CNN) and their application in image classification.	10	L2	CO4
		Module – 5			
Q.9	a.	Define Exploratory Data Analysis (EDA). What is its significance in data science projects?	05	L1	CO5
	b.	Explain the components of time series data: trend, seasonality, and residuals.	05	L2	CO5
	c.	Describe any one real-world case study where Machine Learning is used effectively (e.g., healthcare, finance, etc.).	05	L1	CO5
	d.	Differentiate between Flask and Django frameworks for ML model deployment.	05	L2	CO5
		OR			
Q.10	a.	Define time series data. How is it different from traditional tabular data?	05	L1	CO5
	b.	Compare ARIMA and Prophet models for time series forecasting.	05	L2	CO5
	c.	List and describe any four common data visualization techniques used in EDA.	05	L1	CO5
	d.	What are the key steps in deploying a machine learning model to production?	05	L2	CO5