COURSE TITLE	**MACHINE LEARNING MODELS**
COURSE CODE	**05CA0506**
COURSE CREDITS	**4**

Objective:

1 To familiarize students with core concepts of supervised and unsupervised machine learning using scikit-learn.

2 To train students in building, training, and evaluating regression and classification models.

3 To guide students in selecting appropriate ML models and evaluation metrics for different data problems.

4 To equip students with skills to implement and integrate full ML pipelines from data preprocessing to deployment.

Course Outcomes: After completion of this course, student will be able to:

1 Differentiate between supervised and unsupervised learning approaches and select suitable models for given problems.

2 Implement various machine learning algorithms (e.g., Linear Regression, Decision Trees, KNN, K-Means) using scikit-learn.

3 Evaluate model performance using appropriate metrics such as accuracy, precision, recall, RMSE, and AUC.

4 Design and deploy an end-to-end machine learning project pipeline using real-world datasets.

Pre-requisite of course:NA

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
2	0	4	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	**Advanced EDA Techniques** Multivariate Visualization using Pairplots and Heatmaps, Cross-tabulation and Pivot Tables, Data Transformation – Log, Square Root, Binning, Feature Engineering – Combining, Splitting, Creating new features, Correlation Matrix and Feature Selection, Time Series EDA – Date parsing and plotting, Dealing with Class Imbalance – SMOTE, Undersampling (intro only), Pipelines for Data Preprocessing, Introduction to Interactive EDA using Plotly	15

DR. SUNIL LALCHAND BAJEJA
Digitally signed by (Name of HOD)

DR. RIJWAN KHAN
Digitally signed by (Name of Dean/ Principal)

Contents : Unit	Topics	Contact Hours
2	**Supervised Learning – Regression Models** Introduction to Supervised Learning, Data Splitting – Train/Test and Cross Validation, Linear Regression with scikit-learn, Polynomial Regression & Overfitting, Evaluation Metrics – MSE, RMSE, R^2, Regularization – Ridge and Lasso (basic), Decision Tree Regressor, Random Forest Regressor, Feature Importance in Regression Models	15
Total Hours		**30**

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Unit 1** Supervised Learning, Logistic Regression with scikit-learn, K-Nearest Neighbors (KNN) – Theory & Practice, Decision Trees for Classification, Random Forest Classifier – Concept and Tuning, Evaluation Metrics – Confusion Matrix, Accuracy, Precision, Recall, F1, ROC Curve and AUC, Cross-validation with Stratified K-Fold, Hyperparameter Tuning with GridSearchCV, Saving and Loading Models – Pickle/Joblib, Lab Task: Build and test a Logistics Regression & classification model for spam detection or medical diagnosis.	30
2	**Unit 2** Unsupervised Learning & Capstone Integration, Introduction to Clustering – What and When, K-Means Clustering with Visualization, Elbow Method & Silhouette Score,Hierarchical Clustering & Dendrograms, PCA – Dimensionality Reduction for Visualization, DBSCAN for Density-Based Clustering, Applying Clustering to Customer Segmentation, Creating ML Pipelines (Preprocessing + Model), Project Deployment using Streamlit or Flask (Basic), Capstone Project: Full ML flow from EDA to deployment on a domain-specific dataset	30
Total Hours		**60**

Textbook :

1 Machine Learning: Step-by-Step Guide to Implement Machine Learning Algorithms with Python, Rudolph Russell, Rudolph Russell Publications, 2017

2 Machine Learning, Saikat Dutt, S. Chjandramouli, Das, Pearson, 2019

References:

1 Natural Language Processing with Python, Natural Language Processing with Python, Steven Bird, Ewan Klein, Edward Loper, O'Reilly Media, 2009

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
10.00	15.00	25.00	25.00	25.00	0.00

Instructional Method:

1 Practical, Demo, PPT

Supplementary Resources:

1 https://www.geeksforgeeks.org/machine-learning/

DR. SUNIL LALCHAND BAJEJA
Digitally signed by (Name of HOD)

DR. RIJWAN KHAN
Digitally signed by (Name of Dean/ Principal)