Title of the Course	Machine Learning for Real World Application

Credits	T	P	E	C
	3	1	0	4
Course Type (Theory/Practical/Integrated)	Integrated			
Course Category	Core Discipline			
Pre-Requisite	1. Linear Algebra: matrices, vectors 2. Calculus: differentiation, partial derivatives, gradients 3. Statistics: descriptive statistics, normal distribution, probability 4. Basic syntax and programming constructs 5. Data processing libraries: NumPy, SciPy, Matplotlib, Pandas 6. Basic use of Scikit-learn for machine learning 7. Completion of TCS iON Industry Honour Certification – Artificial Intelligence for Real-World Application			
Learning Objectives	1. Understand the end-to-end machine learning workflow and data engineering practices for real-world problem solving. 2. Apply appropriate preprocessing, feature engineering, and dimensionality reduction techniques to improve model performance. 3. Develop and evaluate supervised learning models for regression and classification tasks. 4. Analyze and implement ensemble and unsupervised learning techniques for complex data scenarios. 5. Explore advanced machine learning concepts including recommender systems, time series forecasting, and reinforcement learning.			

Course Outcomes & Bloom's Level

CO Code	Course Outcome Statement	Bloom Level
CO1	Explain the machine learning lifecycle, CRISP-DM methodology, and advanced data preprocessing techniques.	L2
CO2	Apply feature engineering, scaling, selection, and dimensionality reduction techniques to real-world datasets.	L3
CO3	Build and evaluate regression and classification models using appropriate algorithms and performance metrics.	L3
CO4	Analyze ensemble learning methods and handle imbalanced datasets for improved predictive performance.	L4

	CO5	Evaluate unsupervised learning methods including clustering, anomaly detection, and recommender systems.	L5
	CO6	Design and implement solutions using advanced concepts such as time series forecasting and reinforcement learning.	L6

Course Elements	**Course Element**	**Coverage Level**
	Skill Development	High
	Entrepreneurship	High
	Employability	Moderate
	Professional Ethics	Moderate (Implicit)
	Gender	Low (Implicit)
	Human Values	Low (Contextual)
	Environment & Sustainability	Low (Not Explicit)
SDG (Goals)	SDG 4: Quality Education	
Total Hours of Pedagogy	45 hours Theory (15 hours Self-paced content + 30 hours lecture) 30 hours Practical	

Module#	Content	Pedagogy
M-1	**ML Workflow and Advanced Data Engineering** CRISP-DM Methodology: Detailed steps (Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment); Advanced Data Preprocessing - Handling Missing Data: Imputation strategies (mean, median, mode, regression imputation, MICE), Outlier Detection and Treatment: IQR, Z-score, Isolation Forest, Categorical Encoding: One-Hot, Label, Ordinal, Target Encoding, CatBoost Encoder, Text Vectorization: TF-IDF, CountVectorizer; Feature Engineering: Techniques for creating new features (polynomial features, interaction terms, date/time features, aggregations); Feature Scaling & Normalization: Min-Max, Standardization, Robust Scaling; Feature Selection: Filter, Wrapper (RFE), Embedded methods (Lasso, Tree-based); Dimensionality Reduction: PCA, t-SNE (for visualization).	Self-paced content, Lecture, Modular Assignment
M-2	**Supervised Learning – Regression** Linear Regression: Assumptions, interpretation, multicollinearity; Polynomial Regression and Spline Regression; Regularization: Ridge Regression, Lasso Regression, Elastic Net; Non-Linear Regression Models: Decision Tree Regressor, Random Forest Regressor; Regression Model Evaluation & Diagnostics: MAE, MSE, RMSE, R2, Adjusted R2, residual analysis.	Self-paced content, Lecture, Modular Assignment
M-3	**Supervised Learning – Classification** Logistic Regression: Probability interpretation, multiclass strategies (OvR, OvO); K-Nearest Neighbors (KNN): Distance metrics, computational complexity; Support Vector Machines (SVMs): Maximal margin hyperplane, kernels (Polynomial, RBF, Sigmoid), C-	Self-paced content, Lecture, Modular Assignment

	parameter, gamma; Naïve Bayes Classifiers: Bernoulli, Multinomial, Gaussian; Decision Trees: Gini Impurity, Entropy, pruning techniques; Model Evaluation - Confusion Matrix, Precision, Recall, F1-Score (micro/macro/weighted), ROC Curve, AUC Score, Precision-Recall Curve, Calibration curves; Handling Class Imbalance: SMOTE, ADASYN, Tomek Links, cost-sensitive learning.	
M-4	**Ensemble Methods** Introduction to Semi-supervised Learning: Ensemble Learning, Bias-Variance tradeoff, bagging vs. boosting; Bagging: Bootstrap aggregating, Out-of-Bag (OOB) error; Random Forests: Detailed mechanics, hyperparameter tuning, feature importance; Boosting Algorithms – AdaBoost, Gradient Boosting Machines (GBM), XGBoost: Core concepts (regularization, parallel processing, tree pruning), key parameters, LightGBM & CatBoost: Advantages and use cases; Stacking and Blending: Meta-learners, practical implementation.	Self-paced content, Lecture, Modular Assignment
M-5	**Unsupervised Learning & Advanced ML Concepts** Clustering - K-Means Clustering: Initialization strategies (K-Means++), silhouette score, elbow method, Hierarchical Clustering: Agglomerative vs. Divisive, dendrogram interpretation, DBSCAN: Core points, border points, noise points; Association Rule Mining: Apriori algorithm, support, confidence, lift; Anomaly Detection: Isolation Forest, One-Class SVM, autoencoders for anomaly detection; Time Series Forecasting with ML: Feature engineering for time series, regression models for forecasting, basic Prophet model; Recommender Systems: Collaborative filtering (user-based, item-based), content-based filtering.	Self-paced content, Lecture, Modular Assignment
M-6	**Reinforcement Learning** Foundational Concepts: RL Framework, Markov Decision Process (MDP), Value Functions; Core Algorithms & Techniques: Dynamic programming, Model-free Learning (Monte Carlo Methoda, Temporal-Difference Learning, Q-Learning & SARSA), Exploration-Exploitation Dilemma; Deep Reinforcement Learning: Deep Q-Networks, Policy Gradient Methods; Real-world Applications.	Self-paced content, Lecture, Modular Assignment

List of DIY Modular Assignments

1. Data-Driven Prediction of Customer Purchase Decisions in Retail Finance
2. Interpretable Regression Framework for UK Housing Price Estimation
3. Accurate and Fair Risk Classification
4. Ensemble Learning for Risk and Demand Prediction in Online Retail
5. Unsupervised Learning for Retail Customer Insights
6. Reinforcement Learning for Optimal Energy Policies
7. Responsible AI Audit for Credit Risk Models
8. Designing a Responsible ML Decision-Support System

Modular Assignment Mapping

S. No.	***DIY Assignment Title***	***Mapped CO(s)***	***Bloom Level***
1	Data-driven prediction of customer purchase decisions in retail finance	CO2, CO3	L3
2	Interpretable regression framework for UK housing price estimation	CO3, CO4	L4
3	Accurate and fair risk classification	CO3, CO4, CO5	L5
4	Ensemble learning for risk and demand prediction in online retail	CO4	L4
5	Unsupervised learning for retail customer insights	CO5	L4
6	Reinforcement learning for optimal energy policies	CO6	L6
7	Responsible AI audit for credit risk models	CO5, CO6	L5
8	Designing a responsible ML decision support system	CO6	L6 – Create

Reference Books	1. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition, O'Reilly Media, 2022. 2. Pattern Recognition and Machine Learning, 1st Edition, Springer, 2006. 3. Machine Learning: A Probabilistic Perspective, 1st Edition, MIT Press, 2012. 4. The Elements of Statistical Learning, 2nd Edition, Springer, 2009. 5. Introduction to Machine Learning with Python, 1st Edition, O'Reilly Media, 2016. 6. Machine Learning Yearning, 1st Edition, DeepLearning.AI, 2018.

Course Articulation Matrix

CO \ PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	2	1	2	2	–	–	–	–	–	–	2
CO2	3	3	2	2	3	–	–	–	–	–	–	2
CO3	3	3	3	2	3	–	–	–	–	–	–	2
CO4	2	3	2	3	2	1	–	1	–	–	–	2
CO5	2	3	2	3	2	1	1	2	–	–	–	2
CO6	2	2	3	3	3	2	1	3	1	2	1	3