Advanced ML Techniques										iamneo An NIIT Venture
Day	**Date**	**Track Name**	**Sub Track**	**Topics**	**Subtopics**	**Session 1**	**Session 2**	**Session 3**	**Session 4**	**TOOLS/CASE STUDY**
1	23-12-2025	Supervised Learning	Naive Bayes	Probabilistic modeling, Gaussian Naive Bayes, real-world applications	Gaussian vs Multinomial, applications in classification	Naive Bayes-Overview: Probabilistic classifier based on Bayes' Theorem. Assumptions: Independence of features, Applications: Text classification, spam detection, sentiment analysis.	Gaussian Naive Bayes, Gaussian distribution for continuous data., Assumptions, hyperparameter tuning, and Python implementation.	Multinomial Naive Bayes-Handling discrete data and text features (TF-IDF, Bag of Words). Model training and evaluation (e.g., spam detection).	Real-World Case Study, Problem, preprocessing, model implementation, and tuning., Evaluation and key lessons.	Tools-Scikit-learn, Weka, MATLAB, R (e1071 package), NLTK, TensorFlow, PyTorch, Google Cloud AutoML, RapidMiner, Pandas, Optuna, Spacy.
2	24-12-2025	Supervised Learning	Support Vector Machines (SVM)	Hyperplanes, kernel tricks, tuning SVMs	Linear vs non-linear SVM, soft margin, model evaluation	Overview of SVM, SVM concept: separating hyperplane, Types: Linear and Non-linear SVM	Kernels: Linear, Polynomial, Radial Basis Function (RBF), Concept of margins and support vectors, High-dimensional feature spaces	Tuning SVM parameters, performance metrics - Hyperparameters: C, kernel type, gamma, Grid search and cross-validation, Performance metrics: Accuracy, Precision, Recall, F1-Score	SVM application on text classification - Text preprocessing for SVM, Feature extraction: TF-IDF, Bag of Words, Text classification tasks (e.g., spam detection)	Tools-Scikit-learn, Weka, LIBSVM, MATLAB, TensorFlow, Keras, RapidMiner.
3	26-12-2025	Supervised Learning	Decision Trees	Tree construction, entropy, Gini index	Pruning, overfitting, decision tree visualization	Overview of decision trees, Classification vs Regression trees,How decision trees split data based on features	Understanding Entropy and Gini index - Entropy: Measure of impurity (ID3 algorithm), Gini Index: Impurity measure (CART algorithm), Comparing Entropy and Gini Index	Building and visualizing decision trees - Tree construction using recursive splitting, Visualization tools (e.g., Graphviz, Matplotlib), Interpreting decision trees and feature importance	Overfitting and Pruning in Decision Trees, Overfitting: Trees that fit noise. Pruning: Pre and post-pruning techniques. Avoiding overfitting: Adjusting tree parameters.	Tools-Scikit-learn, Weka, Graphviz, Matplotlib, XGBoost, LightGBM, TensorFlow, RapidMiner.
4	27-12-2025	Model Evaluation	Cross-Validation	K-fold, stratified K-fold, leave-one-out cross-validation	Model comparison methods, model performance metrics	Introduction to Cross-Validation - Overview and purpose (reduces overfitting, improves evaluation), Types: K-Fold, Hold-out, Stratified K-Fold	K-Fold Cross-Validation and implementation - Split data into K folds, train on K-1, validate on 1, Implement using Scikit-learn	Hold-out, stratified K-Fold cross-validation - Hold-out: Train-test split (e.g., 80/20), Stratified K-Fold: Proportional class distribution in each fold	Model performance comparison using cross-validation - Compare models using cross-validation scores (Accuracy, Precision, Recall, F1-Score)	Tools-Scikit-learn, Keras, TensorFlow, Weka, RapidMiner, MATLAB.
5	29-12-2025	Ensemble Learning	Random Forest	Bagging, feature importance, OOB error	Random Forest classifier, model tuning	Introduction to Random Forest - Ensemble method: Multiple decision trees, Benefits: Reduces overfitting, improves accuracy	Understanding OOB error and feature importance - OOB Error: Estimate using unused data, Feature Importance: Measures feature contribution	Training Random Forest model - Build multiple trees, aggregate results, Key parameters: Number of trees, max depth	Model evaluation using Random Forest - Metrics: Accuracy, Precision, Recall, F1-Score, Hyperparameter tuning and cross-validation	Tools-Scikit-learn, Weka, MATLAB, XGBoost, LightGBM, RapidMiner, H2O.ai.
6	30-12-2025	Ensemble Learning	Boosting Techniques	AdaBoost, boosting methods	Model stacking, XGBoost integration	Overview of AdaBoost and boosting concept, How boosting improves weak learners	XGBoost implementation, hyperparameter tuning - Implementing XGBoost for classification/regression, Key hyperparameters: Learning rate, max depth, n_estimators	Boosting models for classification tasks - Using boosting algorithms for classification (e.g., AdaBoost, XGBoost), Handling imbalanced data with boosting	Model performance comparison using Boosting - Comparing boosting models with other algorithms (e.g., SVM, Random Forest), Performance metrics: Accuracy, Precision, Recall, F1-Score	Tools-Scikit-learn, XGBoost, LightGBM, AdaBoost, CatBoost, H2O.ai, RapidMiner.
7	19-12-2025	Unsupervised Learning	Hierarchical Clustering	Dendrograms, agglomerative vs divisive	Linkage criteria, cluster interpretation	Overview of hierarchical clustering, Types: Agglomerative and Divisive	Dendrogram: Tree-like diagram to visualize clustering, Agglomerative: Bottom-up approach for clustering	Divisive: Top-down approach for clustering, Linkage criteria: Single, Complete, Average, Ward's method	Cluster interpretation using Hierarchical Clustering - Interpreting dendrogram to define clusters, Selecting optimal number of clusters	Tools-Scikit-learn, SciPy, Weka, MATLAB, RapidMiner, H2O.ai.

8	20-12-2025	Unsupervised Learning	DBSCAN Clustering	Density-based clustering, epsilon-neighborhood	Core points, noise, tuning parameters	Introduction to DBSCAN - Density-based clustering method. Identifies core, border, and noise points.	Epsilon-neighborhood and density-based clustering - Epsilon: Defines neighborhood range. Clustering based on density of points.	Handling noise/outliers in DBSCAN - Detects and handles noise points/outliers.	Parameter tuning for DBSCAN and cluster visualization - Key parameters: Epsilon (ε) and MinPts. Visualizing clusters and tuning parameters.	Tools-Scikit-learn, DBSCAN, Weka, MATLAB, H2O.ai, RapidMiner, PyClustering.
9	21-12-2025	Dimensionality Reduction	PCA & LDA	PCA for variance preservation, LDA for class separation	Dimensionality reduction techniques	Overview of Principal Component Analysis (PCA),Dimensionality reduction technique	PCA for variance preservation - Maximizing variance with principal components, Choosing components based on explained variance	LDA for class separation - Linear Discriminant Analysis (LDA) for supervised dimensionality reduction, Enhancing class separability by maximizing between-class variance	Visualization of transformed data using PCA and LDA - Visualizing data in reduced dimensions, Plotting PCA and LDA components for better understanding	Tools-Scikit-learn, MATLAB, Weka, PCA, LDA (via SciPy), TensorFlow, H2O.ai, RapidMiner.
10	22-12-2025	GenAI Tools	ML Workflow Automation	GenAI-powered model recommendation, AutoML	Smart feature engineering, automation of ML workflows	Introduction to GenAI Tools - Overview and applications of GenAI tools.	AutoML pipeline integration - Automating machine learning workflows (data, model selection, evaluation).	Feature engineering and model automation - Automating feature engineering and model training.	Final project: Workflow automation using GenAI tools - Designing end-to-end workflow automation with GenAI tools.	Tools-Google Cloud AutoML, H2O.ai, Microsoft Azure AutoML, Amazon SageMaker, RapidMiner, DataRobot, MLflow.
11	23-12-2025	**REVISION**								
12	24-12-2025	**FINAL END SEMESTER EXAMINATION**								