

Machine Learning

★ MODULE 1: Introduction to Machine Learning

- What is ML? AI vs ML vs DL vs Data Science
 - Types of ML: Supervised, Unsupervised, Semi-supervised
 - Real-world ML applications
 - ML project lifecycle: Data → Model → Evaluation
 - Python ecosystem for ML (Jupyter, Colab, Scikit-learn)
 - Understanding datasets, labels, features
-

★ MODULE 2: Python Essentials for Machine Learning

✓ Python Basics (quick revision)

- Variables, lists, dictionaries
 - Loops & conditions
 - Functions
 - File handling
 - Working with Jupyter Notebook & Google Colab
-

✓ NumPy (Numerical Computing)

- NumPy arrays
 - Array indexing & slicing
 - Mathematical operations
 - Broadcasting
 - Aggregations (mean, sum, std)
-

✓ Pandas (Data Manipulation)

- DataFrames & Series
 - Reading CSV/Excel/JSON
 - Filtering, sorting, grouping
 - Handling missing values
 - Merging, joining, concatenating
 - Feature engineering basics
 - Date-time operations
 - Handling duplicates
 - Descriptive statistics
-

✓ Matplotlib & Seaborn (Data Visualization)

- Line, bar, pie, histogram, scatter plots
 - Heatmaps & pairplots
 - Outlier visualization
 - Correlation matrix
 - Trend analysis
-

✓ Basic Exploratory Data Analysis (EDA)

- Understanding dataset structure
 - Data types, distributions
 - Outlier detection
 - Correlation analysis
 - Insights for business problems
-

Projects

- EDA on Titanic dataset
 - EDA on Sales dataset
 - EDA on Iris dataset
-

★ MODULE 3: Math & Statistics for ML

- Linear Algebra: vectors, matrices, dot product
 - Probability: Bayes theorem, distributions
 - Statistics: mean, variance, std, skew, correlations
 - Gradient Descent (intuitive explanation)
 - Cost functions for regression/classification
 - Evaluation Metrics Overview
-

★ MODULE 4: Data Preprocessing & Feature Engineering

- Handling missing values, outliers
 - Data cleaning & transformation
 - Categorical encoding: Label, One-Hot Encoding
 - Scaling: Standardization, Normalization
 - Train-test split, Cross-validation
 - Feature selection: Filter, Wrapper, Embedded
 - Feature extraction: PCA
 - Data pipeline building (Scikit-learn Pipeline)
-

★ MODULE 5: Supervised Learning — Regression

- Linear Regression
 - Multiple Regression
 - Polynomial Regression
 - Ridge, Lasso, ElasticNet
 - Evaluation metrics: MAE, MSE, RMSE, R²
 - Residual analysis
-

Mini Projects

- House price prediction
 - Salary prediction
 - Sales forecasting (basic)
-

★ MODULE 6: Supervised Learning — Classification

- Logistic Regression
 - KNN Classifier
 - Decision Trees
 - Random Forest
 - Naive Bayes
 - SVM (Support Vector Machines)
 - Class imbalance techniques (SMOTE)
 - Evaluation metrics:
 - Accuracy, Precision, Recall
 - F1, ROC-AUC, Confusion Matrix
 - Precision-Recall Curve
-

Mini Projects

- Spam detection
 - Fraud detection
 - Loan approval classification
-

★ **MODULE 7: Ensemble Methods & Advanced ML**

Models

- Bagging, Boosting, Stacking
- AdaBoost
- Gradient Boosting
- XGBoost, LightGBM, CatBoost
- Feature importance analysis
- Model interpretability (SHAP, LIME)

Projects

- Telecom churn prediction
- Credit risk scoring

★ **MODULE 8: Unsupervised Learning**

- Clustering: K-Means, Hierarchical, DBSCAN
- Dimensionality reduction: PCA, t-SNE (intro)
- Anomaly detection
- Association Rule Learning: Apriori, Eclat

Projects

- Customer segmentation
- Market basket analysis
- Anomaly detection in transactions

★ **MODULE 9: Time Series Forecasting**

- Time series decomposition
- Stationarity, ADF test
- Moving averages
- ARIMA basics
- SARIMA basics
- Facebook Prophet (2025 trending tool)

Projects

- Sales forecasting
- Demand forecasting
- Time-based predictions

★ **MODULE 10: Natural Language Processing (NLP) — Classical Methods**

- Text cleaning: tokenization, stopwords, lemmatization
- Bag of Words
- TF-IDF
- N-grams
- Word embeddings (Word2Vec intro)
- Text classification
- Sentiment analysis

Projects

- Movie review sentiment analysis
- News category classification

★ **MODULE 11: Intro to Deep Learning**

★ **MODULE 12: Intro to Computer Vision**

- Image processing basics
- Flattened pixel features
- Basic CNN (1–2 layers)

- Transfer learning overview (ResNet/VGG introduction only)

Projects

- Simple image classifier
- Digit recognition

★ **MODULE 13: Model Validation, Optimization &**

Best Practices

- Train-validation-test strategy
- Bias-variance tradeoff
- Hyperparameter tuning:
 - GridSearchCV
 - RandomizedSearchCV
- Feature importance
- Reducing overfitting
- Building reusable ML pipelines

★ **MODULE 13: Model Deployment (Optional)**

- Saving models (Pickle, Joblib)
- Deployment using:
 - Flask
 - FastAPI
 - Streamlit / Gradio
- Hosting:
 - Localhost
 - Render / HuggingFace Spaces
 - AWS EC2 (basic)
- Building simple ML web apps

Projects

- House price ML web app
- Sentiment analysis UI
- Fraud detection web API

★ **MODULE 14: Capstone Projects**

1 Regression Project

2 Classification Project

3 NLP or Time Series Project

4 Optional Image-based ANN Project

★ **Tools & Technologies Used**

Programming: Python

Libraries: Numpy, Pandas, Matplotlib, Seaborn

ML: Scikit-learn, XGBoost, CatBoost, LightGBM

DL Intro: TensorFlow/Keras

NLP: NLTK, SpaCy

Visualization: Seaborn, Plotly

Deployment: Streamlit, Flask, FastAPI

Versioning: Git, GitHub