Mastering Machine Learning with Python

Duration: 10 days
Prerequisites: Knowledge of Python Programming and Machine Learning

Day 1 – Python Essentials for Machine Learning & Environment Setup

Topics:

- Introduction to AI, ML, DL, Data Science ecosystem
- Machine Learning lifecycle and workflow
- Python refresher: data types, loops, functions, file handling
- Working with libraries: **NumPy, Pandas, Matplotlib, Seaborn**
- Understanding arrays, broadcasting, indexing, slicing
- Vectorization and matrix operations
- Working with Jupyter Notebooks and virtual environments
- Overview of datasets: CSV, JSON, APIs

Labs:

- Set up Anaconda environment and Jupyter Notebook
- Import data using Pandas and perform operations
- Calculate descriptive statistics and visual summaries

Day 2 – Data Wrangling, Cleaning, and Feature Engineering

Topics:

- Handling missing, duplicate, and inconsistent data
- Data imputation strategies (mean, mode, median, predictive)
- Encoding categorical variables (One-Hot, Label, Ordinal)
- Feature scaling (Standard, MinMax, Robust, Log scaling)
- Binning continuous data and feature discretization
- Outlier detection (IQR, Z-Score, Isolation Forest)
- Feature extraction and transformation
- Handling date/time and text features

Labs:

- Clean and preprocess the **Titanic** dataset end-to-end
- Encode and scale features; visualize pre/post results
- Create new features (feature interaction, polynomial features)

Day 3 – Exploratory Data Analysis (EDA) and Visualization

Topics:

- Data profiling and summary statistics
- Correlation, covariance, and hypothesis testing

- Skewness, kurtosis, and normality checks
- Pair plots, boxplots, histograms, violin plots
- Multi-feature visualization with Seaborn & Plotly
- Feature selection using variance, correlation, and ANOVA
- Detecting multicollinearity (VIF)
- Creating data storytelling reports

Labs:

- EDA on **E-Commerce Sales Dataset**
- Identify correlations and visualize trends
- Generate automated profiling using **pandas-profiling / ydata-profiling**

Day 4 – Regression Models (Supervised Learning – Part 1)

Topics:

- Concept of regression, bias-variance trade-off
- Linear Regression, Polynomial Regression
- Regularization: Ridge, Lasso, ElasticNet
- Evaluation metrics: MAE, MSE, RMSE, R^2
- Residual analysis and model diagnostics
- Gradient Descent and Cost Function intuition
- Train/Test Split, Cross-Validation

Labs:

- Predict house prices using Boston dataset
- Compare Lasso, Ridge, and ElasticNet performance
- Visualize loss convergence and coefficient shrinkage

Day 5 – Classification Models (Supervised Learning – Part 2)

Topics:

- Logistic Regression fundamentals
- K-Nearest Neighbors, Decision Trees, Random Forest
- Support Vector Machines (linear and kernel)
- Naïve Bayes (Gaussian, Multinomial)
- Performance metrics: Accuracy, F1, Precision-Recall, ROC-AUC
- Handling imbalanced data (SMOTE, class weighting)
- Model calibration and threshold tuning

Labs:

- Classify **heart disease** and **Titanic survivors**
- Visualize decision boundaries for SVM/KNN
- Evaluate models using confusion matrix and ROC curves

Day 6 – Unsupervised Learning & Dimensionality Reduction

Topics:

- Concept of clustering and similarity metrics
- K-Means, Hierarchical, DBSCAN clustering
- Choosing optimal K (Elbow, Silhouette)
- Principal Component Analysis (PCA), t-SNE, UMAP
- Feature compression and visualization
- Association Rule Mining (Apriori, FP-Growth)
- Anomaly detection with Isolation Forest

Labs:

- Cluster customer data using K-Means
- Visualize clusters and apply PCA reduction
- Perform association analysis on retail transaction dataset

Day 7 – Ensemble & Advanced ML Algorithms

Topics:

- Ensemble Learning fundamentals: Bagging vs Boosting
- Random Forest, AdaBoost, Gradient Boosting
- XGBoost, LightGBM, CatBoost comparison
- Stacking and Blending models
- Hyperparameter tuning (GridSearchCV, RandomizedSearchCV, Optuna)
- Cross-Validation, K-Fold, StratifiedKFold
- Feature importance visualization

Labs:

- Customer churn prediction using XGBoost
- Tune hyperparameters with GridSearchCV
- Compare ensemble models' accuracy

Day 8 – Model Optimization, Pipelines, and Automation

Topics:

- Model overfitting vs underfitting
- Regularization and dropout techniques
- ML pipelines using scikit-learn's Pipeline
- Saving and loading models (Pickle, Joblib)
- AutoML tools overview (PyCaret, Auto-Sklearn)
- Model explainability with LIME & SHAP
- Handling large datasets and parallelization

Labs:

- Build and automate ML pipeline for loan prediction
- Explain model outputs using SHAP/LIME
- Run quick experiments with PyCaret

Day 9 – Deep Learning & Neural Networks

Topics:

- Neural Network architecture: layers, weights, activation functions
- Forward & backward propagation, loss, and optimization
- Introduction to **TensorFlow** and **Keras**
- Building Feedforward Neural Network (ANN)
- Image classification with CNN (overview)
- Overfitting prevention (dropout, regularization)
- Transfer learning concepts

Labs:

- Build ANN for predicting customer churn
- Train and evaluate model using Keras
- Visualize learning curves and confusion matrix

Day 10 – Model Deployment, MLOps, and Capstone

Topics:

- Model serialization (Pickle/Joblib)
- Deploy ML model using **Flask** and **Streamlit**
- Integrating models with REST APIs
- Docker basics for ML deployments
- Introduction to MLOps (CI/CD, model monitoring)
- Ethics in AI: Bias, fairness, and explainability
- Capstone Project Presentations

Labs:

- Create and deploy model with Streamlit UI
- Test endpoint via REST API
- Present final project and get peer feedback