MACHINE LEARNING BASICS TRACK – 2025 EDITION

" From zero to your first ML model."

Market Demand Note

Machine Learning (ML) is one of the most in-demand tech skills globally, powering applications from recommendations to fraud detection. Even a basic understanding of ML concepts can open career opportunities in data science, AI, and analytics. This program gives participants hands-on exposure to ML fundamentals.

Duration: 6 weeks | **Mode**: Online/Offline

📘 Machine Learning basics Track — 2025 Edition

Table of Contents

Week 6: Final Project – Predicting House Prices

- End-to-end ML project
- Data cleaning, feature engineering, model training, evaluation, visualization
- Hands-on: Deliver a Jupyter Notebook with code, visualizations, and insights

Detailed Content

Week 1: Introduction to Machine Learning

- **Understand what Machine Learning (ML) is**
  (Computers learn from data without explicit programming; examples: spam detection, recommendation systems)
- **Explore real-world ML applications**
  (Healthcare, finance, retail, automation)
- **Learn the three main types of ML:**
  - **Supervised Learning (labeled data, e.g., regression, classification)**
  - **Unsupervised Learning (unlabeled data, e.g., clustering)**
  - **Reinforcement Learning (learning by interaction, e.g., robotics, games)**
- **Set up Python, Jupyter Notebook, and essential libraries**
  (Install Anaconda/Miniconda, create your first notebook)
- **Hands-on:**
  - **Install Python and Jupyter Notebook**
  - **Load a sample dataset (e.g., iris, diabetes)**
  - **Write and execute your first Python code in Jupyter**

Week 2: Data Preparation

- **Load datasets from CSV and Excel files using Pandas**
- **Inspect data: head, tail, info, describe**
- **Handle missing values: fill, drop, impute**
- **Detect and handle outliers: visual and statistical methods**
- **Perform basic feature engineering: create new features, transform existing ones**

- **Hands-on:**
    - **Import and clean a real-world dataset (e.g., housing, sales)**
    - **Practice handling missing data and outliers**
    - **Create and document feature engineering steps**

Week 3: Supervised Learning – Part 1

- **Linear Regression:**
    - **Understand the concept of fitting a line to data**
    - **Implement using scikit-learn**
    - **Interpret coefficients and predictions**
- **Logistic Regression (Classification):**
    - **Binary classification basics**
    - **Implement using scikit-learn**
    - **Understand decision boundaries**
- **Model Evaluation:**
    - **Accuracy, precision, recall (for classification)**
    - **Mean squared error, R^2 (for regression)**
- **Hands-on:**
    - **Train and evaluate a linear regression model**
    - **Train and evaluate a logistic regression model**
    - **Compare model performance using metrics**

Week 4: Unsupervised Learning

- **Clustering with K-Means:**
    - **Group similar data points**
    - **Choose the number of clusters (elbow method)**
    - **Interpret cluster results**
- **Dimensionality Reduction with PCA:**
    - **Reduce the number of features while preserving variance**
    - **Visualize high-dimensional data in 2D/3D**

- **Hands-on:**
    - **Apply K-Means to a dataset (e.g., customer segmentation)**
    - **Reduce dimensions using PCA and visualize the results**

Week 5: Improving ML Models

- **Train/Test Split:**
    - **Why separate training and test data**
    - **Avoid data leakage**
- **Cross-Validation:**
    - **k-fold cross-validation for robust evaluation**
- **Avoid Overfitting:**
    - **Understand bias-variance tradeoff**
    - **Regularization basics**
- **Feature Scaling:**
    - **Standardization (Z-score normalization)**
    - **Normalization (Min-Max scaling)**
- **Hands-on:**
    - **Implement and compare train/test split vs. cross-validation**
    - **Apply feature scaling and observe impact on model performance**
    - **Discuss best practices to avoid overfitting**

Week 6: Final Project – Predicting House Prices

- **Understand the problem: Predict house prices using historical data**
- **Data Preparation:**
    - **Load, clean, and explore the dataset**
    - **Perform feature engineering**
- **Model Training:**
    - **Choose and train regression models (e.g., linear regression, decision tree)**
    - **Evaluate using appropriate metrics**

- **Visualization:**
    - **Plot actual vs. predicted values**
    - **Visualize feature importances**
- **Documentation:**
    - **Write clear comments and markdown in Jupyter Notebook**
    - **Present findings and insights**
- **Hands-on:**
    - **Deliver a complete Jupyter Notebook with code, visualizations, and explanations**
    - **Present your project to peers, explain your approach and results**

Tools Covered

- **Python (core programming)**
- **scikit-learn (ML algorithms)**
- **Pandas (data manipulation)**
- **NumPy (numerical operations)**
- **Matplotlib & Seaborn (data visualization)**
- **Jupyter Notebook (interactive coding and reporting)**

Final Project Guidelines

Title: Predicting House Prices
Goal: Build a complete ML pipeline from raw data to predictions
Tasks:

- **Load and clean the dataset**
- **Perform exploratory data analysis and feature engineering**
- **Train and evaluate ML models**
- **Visualize results and document your process**
  Output: Jupyter Notebook with working code, visualizations, and a clear explanation of each step

Assessment & Certification

- **Weekly assignments: Labs, quizzes, and mini projects.**
- **Final project: End-to-end ML solution with code, visualizations, and presentation.**
- **Certification: Awarded upon successful completion of all assignments and final project.**