Foundations of Machine Learning

By

Dr Ravi Prakash Verma
Professor

Department of CSAI

ABESIT

Glimpse

- **Total Duration: 39+ Hours (13 Days)**
- **Mode: Hands-on coding in Python (NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn).**
- **Target Audience: Students**
- **Prerequisites: Python programming , pandas, mathplot, numpy and fundamental statistics knowledge.**

Objectives

- **By the end of this workshop, participants will:**

  ✅ **Understand theoretical concepts behind key ML models.**

  ✅ **Implement models using Python and Scikit-learn.**

  ✅ **Analyze real-world datasets for predictive modeling.**

  ✅ **Optimize models using feature selection and hyperparameter tuning.**

  ✅ **Evaluate models using appropriate performance metrics (R^2, RMSE, MAE, ROC-AUC, etc.).**

Workshop Structure & Duration

Day	Topics Covered
Day 1	Introduction to ML & Data Preprocessing
Day 2	Linear Regression - Concept & Mathematics
Day 3	Linear Regression in Python
Day 4	Model Performance Metrics (R^2, RMSE, MAE, MSE)
Day 5	Confusion Matrix & Model Performance
Day 6	Logistic Regression - Concept & Mathematics
Day 7	Project on Logistic Regression in Python

Workshop Structure & Duration

Day 8	**Decision Trees - Concept & Theory**
Day 9	Project on Decision Trees in Python
Day 10	Support Vector Machines (SVM) - Linear SVM
Day 11	Support Vector Machines (SVM) - Non-Linear SVM
Day 12	Project on Linear & Non-Linear SVM in Python
Day 13	Future Learning Path (Deep Learning, NLP, etc.)

Detailed Workshop Contents

- **Day 1: Introduction & Data Preprocessing (3 Hours)**
- **Theory:**
  - **Overview of Supervised vs. Unsupervised Learning**
  - **Importance of Feature Engineering**
  - **Handling missing values, outliers, and categorical variables**
  - **Data normalization & standardization**
- **Hands-on Coding:**
  - **Importing datasets using Pandas**
  - **Handling missing data using mean/median imputation**
- **Feature scaling using MinMaxScaler & StandardScaler**

Detailed Workshop Contents

- **Day 2: Linear Regression - Concept & Mathematics (3 Hours)**
- **Theory:**
  - **Introduction to Regression Problems**
  - **Mathematical formulation of Simple & Multiple Linear Regression**
  - **Understanding Cost Function & Gradient Descent**
- **Hands-on Coding:**
  - **Implementing Linear Regression from scratch using NumPy**
  - **Visualizing best-fit line using Matplotlib**

Detailed Workshop Contents

- **Day 3: Linear Regression in Python (3 Hours)**
- **Theory:**
  - **Feature Selection & Model Complexity**
  - **Handling Multicollinearity using VIF**
- **Hands-on Coding:**
  - **Using Scikit-learn's LinearRegression**
- **Implementing Feature Selection Techniques**

Detailed Workshop Contents

- **Day 4: Model Performance Metrics (R^2, RMSE, MAE, MSE) (3 Hours)**
- **Theory:**
- **Understanding R^2 (Coefficient of Determination)**
- **Root Mean Squared Error (RMSE) vs. Mean Absolute Error (MAE)**
- **Importance of Adjusted R^2 for multiple regression**
- **Hands-on Coding:**
- **Implementing R^2, RMSE, MAE, and MSE in Python**
- **Comparing models using different metrics**
- **Confusion Matrix & Model Performance**
  - **True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN)**
  - **Precision, Recall, and F1-Score**
  - **Visual Representation of Confusion Matrix**
  - **Code Implementation for Confusion Matrix**
- **ROC & AUC Score**
  - **Understanding ROC Curve**
  - **Computing AUC Score for Model Performance**

Detailed Workshop Contents

- **Day 5: Confusion Matrix & Model Performance**

- **Confusion Matrix & Model Performance**
  - **True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN)**
  - **Precision, Recall, and F1-Score**
  - **Visual Representation of Confusion Matrix**
  - **Code Implementation for Confusion Matrix**
- **ROC & AUC Score**
  - **Understanding ROC Curve**
  - **Computing AUC Score for Model Performance**

Detailed Workshop Contents

- **Day 6: Logistic Regression - Concept & Mathematics (3 Hours)**
- **Theory:**
  - **Difference between Regression & Classification**
  - **Sigmoid Function & Decision Boundary**
- **Understanding Cost Function for Logistic Regression**
- **Hands-on Coding:**
  - **Implementing Sigmoid Function from scratch**
  - **Plotting decision boundaries for classification**

Detailed Workshop Contents

- **Day 7: Project on Logistic Regression in Python (3 Hours)**
- **Theory:**
  - **Confusion Matrix, Precision, Recall, F1-Score**
  - **Introduction to ROC & AUC for classification**
- **Hands-on Coding:**
  - **Implementing Logistic Regression using Scikit-learn**
- **Plotting ROC Curve & AUC Score**

Detailed Workshop Contents

- **Day 8: Decision Trees - Concept & Theory (3 Hours)**
- **Theory:**
  - Understanding Entropy, Information Gain, and Gini Impurity
  - Overfitting in Decision Trees
  - Pruning Techniques
- **Hands-on Coding:**
  - Implementing Decision Tree from scratch
  - Visualizing Tree Structure using Graphviz

Detailed Workshop Contents

- **Day 9: Project on Decision Trees in Python (3 Hours)**
- **Theory:**
  - **Feature Importance in Decision Trees**
  - **Hyperparameter Tuning using GridSearchCV**
- **Hands-on Coding:**
  - **Implementing DecisionTreeClassifier using Scikit-learn**
  - **Optimizing decision trees for better accuracy**

Detailed Workshop Contents

- **Day 10: Linear SVM - Concept & Theory (3 Hours)**
- **Theory:**
  - **Introduction to Support Vector Machines (SVM)**
  - **Understanding Hyperplanes & Support Vectors**
  - **Hard Margin vs. Soft Margin SVM**
- **Hands-on Coding:**
  - **Implementing Linear SVM using Scikit-learn**
  - **Visualizing decision boundaries in 2D classification problems**

Detailed Workshop Contents

- **Day 11: Non-Linear SVM - Concept & Theory (3 Hours)**
- **Theory:**
  - Understanding Kernel Trick for Non-Linear Classification
  - Types of Kernels: Polynomial, Gaussian (RBF), Sigmoid
  - Tuning SVM Parameters (C, Gamma, Kernel Choice)
- **Hands-on Coding:**
  - Implementing Non-Linear SVM using RBF Kernel
  - Visualizing Complex Decision Boundaries

Detailed Workshop Contents

- **Day 12: Project on Linear & Non-Linear SVM in Python (3 Hours)**
- **Theory:**
  - **Comparing Linear vs. Non-Linear SVM Performance**
  - **Choosing the Right Kernel for Different Datasets**
- **Hands-on Coding:**
  - **Training Linear and Non-Linear SVM models**
  - **Hyperparameter tuning using GridSearchCV**

Detailed Workshop Contents

- **Day 13: Future Learning Path**
- Theoretical discussion on
  - **(Deep Learning, NLP, etc.)**

Learning Outcomes

- **By the end of this workshop, participants will:**

  ✅ **Master Linear Regression, Logistic Regression, Decision Trees, and SVM.**

  ✅ **Implement both Linear & Non-Linear SVM with different kernels.**

  ✅ **Optimize models using feature selection & hyperparameter tuning.**

  ✅ **Evaluate models using R^2, RMSE, MAE, ROC-AUC, and more.**