# 📚 Machine Learning Basics — FULL REVISION

---

## ✅ Day 1: Introduction to Supervised Learning

### 🔹 What is Supervised Learning?

- You train the model on **labeled data** (input + correct output).
- Two types:
    - **Regression** → Predict numbers (e.g., price)
    - **Classification** → Predict categories (e.g., spam/not spam)

### 🔍 Real-life Example:

- Regression: Predicting house price from size
- Classification: Predicting if tumor is benign or malignant

📁 *Notebook:* `Supervised_learning_intro.ipynb`

---

## ✅ Day 2: Linear Regression

### 🔹 Simple Linear Regression:

- One input feature (X) → one output (Y)
- Formula: `y = mx + c`
  (m = slope, c = intercept)

### 🔹 Multiple Linear Regression:

- More than one feature: `y = m1x1 + m2x2 + ... + c`

### 📊 You learned:

- How to plot the regression line
- Evaluate using **R² score** (closer to 1 = better)
- Residual plots to check assumptions

📁 *Notebook:* `Linear_Regression_in_depth.ipynb`

---

# ✅ Day 3: Logistic Regression

### 🔹 Goal:

- Predict probabilities for classification problems (output is 0 or 1)

### 📌 You implemented:

- Logistic regression on **Breast Cancer dataset**
- Evaluated with:
    - **Confusion Matrix** (TP, FP, FN, TN)
    - **ROC Curve** and **AUC Score**
    - **Accuracy, Precision, Recall, F1-Score**

📁 *Notebook:* `Logistic_Regression_Breast_Cancer_data.ipynb`

---

# ✅ Day 4: Decision Trees & Random Forests

### 🌳 Decision Trees:

- Splits data into branches using feature thresholds
- Good for both classification & regression

### 🌲 Random Forest:

- Multiple decision trees combined (ensemble)
- More accurate and robust
- Feature importance gives you insight into which features matter most

📁 *Folder:* `Decision Trees & Random Forests`

---

## ✅ Day 5 (Part A): Model Evaluation Metrics

### 🔹 Classification Metrics:

- **Accuracy**: Correct predictions / total
- **Precision**: Correct positives out of predicted positives
- **Recall**: Correct positives out of actual positives
- **F1 Score**: Harmonic mean of Precision & Recall
- **ROC Curve**: True Positive Rate vs. False Positive Rate
- **AUC**: Area under ROC Curve (1 = best)

---

## ✅ Day 5 (Part B): Unsupervised Learning

### 📍 K-Means Clustering:

- Groups data into K clusters
- You used **Elbow Method** to choose the best K
- Visualized clusters

### 📍 PCA (Principal Component Analysis):

- Reduced number of features
- Kept only the most important directions of variance
- Used for visualization and speeding up models

📁 *Folder:* `Model Evaluation & Unsupervised Learning`

---

## ✅ Day 6 (Part A): Model Tuning — Data Split

### 🔹 `train_test_split`:

- Split your data into training and testing sets (usually 80/20)

### 🔹 Avoid Data Leakage:

- Don't leak test data info into training
- Always split first, then scale or encode separately

---

## ✅ Day 6 (Part B): Cross Validation + GridSearchCV

### 🔹 K-Fold Cross Validation:

- Divides data into K parts and rotates training/testing

- Reduces overfitting, gives better model reliability

### 🔹 GridSearchCV:

- Searches over a grid of hyperparameters
- Picks the best model using cross-validation scores

📁 *Folder:* `Classification Model Tuning`

---

# ✅ Day 7: ML Mini Project — House Price Prediction

### 📂 You Did:

- Cleaned & preprocessed the dataset (nulls, scaling, encoding)
- Applied multiple models:
    - Linear Regression
    - Decision Tree
    - Random Forest
    - Ridge & Lasso (regularized regressions)

### 📊 Evaluated Using:

- **MAE**: Mean Absolute Error
- **MSE**: Mean Squared Error
- **RMSE**: Root Mean Squared Error
- **R² Score**: Measures goodness of fit

📁 *Project:* `House_Price_Prediction_ML`

---

## 🔗 GitHub Repo Summary

- 📁 Organized folders for each day's topic
- ✅ Well-documented notebooks
- 📸 Visualizations of models and performance
- 💡 Great for showing recruiters your hands-on skills

---

## 🎯 What You Mastered

| Skill | Tools/Concepts |
|---|---|
| Regression & Classification | Linear & Logistic Regression |
| Tree-based Models | Decision Tree, Random Forest |
| Model Evaluation | Accuracy, Precision, Recall, ROC-AUC |
| Unsupervised Learning | K-Means, PCA, Clustering, Elbow method |
| Model Tuning | Cross-validation, GridSearchCV |
| Real-world ML project | Data Cleaning, Feature Engineering, Multiple Models |