# MNIST Digit Classification using Traditional ML Algorithms

## 1. Project Overview

This project classifies handwritten digits (0–9) using traditional machine learning models. The dataset used is the classic MNIST dataset from Kaggle, which contains grayscale images of handwritten digits. The goal is to compare the performance of various ML models based on F1-score.

## 2. Data Handling & Feature Engineering

Images were extracted from a compressed zip file and processed into pixel arrays. Each image was flattened into a feature vector suitable for ML model input. Preprocessing included standard normalization and data splitting for training/testing.

## 3. Machine Learning Models Used

- Logistic Regression  
- Decision Tree  
- Random Forest  
- K-Nearest Neighbors (KNN)  
- Gaussian Naive Bayes  
- Support Vector Machine (SVM)  
- AdaBoost  
- XGBoost

## 4. Results & Evaluation

Model performance was evaluated using F1-score. Among the models tested, XGBoost achieved the highest F1-score of 0.989, followed closely by Random Forest (0.986) and SVM (0.978).

## 5. Conclusion & Next Steps

The project demonstrates that even without deep learning, traditional ML algorithms can achieve high accuracy on digit classification tasks. Future improvements may include PCA for dimensionality reduction, hyperparameter tuning, or transitioning to CNNs for deeper insights.