

# Hand Digit Recognition Using ML

---

## Introduction

The project focuses on digit classification using the MNIST dataset. The MNIST dataset comprises handwritten digits (0-9), and our goal is to build a machine learning model that accurately classifies these digits. This project demonstrates the application of logistic regression for image classification, which is a fundamental task in computer vision.

## Data Description

- **Dataset:** The dataset used is the MNIST digits dataset, which is a standard benchmark in the field of machine learning.
- **Features:** The dataset consists of 8x8 pixel images of handwritten digits.
- **Target:** The target labels are the digits (0-9).

## Steps and Methodology

1. **Importing Libraries and Dataset:**
  - Imported necessary libraries such as numpy, pandas, seaborn, matplotlib, and scikit-learn.
  - Loaded the MNIST digits dataset from `sklearn.datasets`.
2. **Exploratory Data Analysis (EDA):**
  - Visualized the first 16 images of the dataset along with their labels to understand the data.
3. **Data Preprocessing:**
  - Reshaped the image data from 8x8 pixel images to a flat array of 64 features.
  - Applied Min-Max scaling to normalize the data, ensuring that all pixel values are between 0 and 1.
4. **Data Splitting:**
  - Split the data into training and testing sets using an 75-25% split.
5. **Model Training:**
  - Trained a Logistic Regression model using the training data.
6. **Model Evaluation:**
  - Made predictions on the test data.
  - Evaluated the model using classification metrics such as classification report, accuracy score, and confusion matrix.

## Results

1. **Classification Report:**
  - The classification report provides precision, recall, and F1-score for each digit class.
1. **Accuracy Score:**
  - The overall accuracy of the model is approximately 94%.
2. **Confusion Matrix:**

- The confusion matrix visualizes the performance of the classifier, showing true positives, false positives, true negatives, and false negatives for each class.

## Conclusion

- The logistic regression model performed well, achieving an accuracy of 94% on the test set.
- The model is effective in classifying handwritten digits, making few errors.
- The confusion matrix indicates that most of the misclassifications are close to the true digit, which suggests the model's high confidence in predictions.

## Future Work

- **Model Improvement:** Consider using more advanced models like Convolutional Neural Networks (CNNs) for better accuracy and performance.
- **Hyperparameter Tuning:** Experiment with hyperparameter tuning for the logistic regression model to further improve performance.
- **Data Augmentation:** Apply data augmentation techniques to increase the diversity of the training set, potentially improving the model's robustness.
- **Feature Engineering:** Explore additional feature engineering techniques to enhance model performance.
- **Deploying the Model:** Develop a web application or mobile app to deploy the model, allowing users to upload images of handwritten digits and receive predictions.