

Organization: Elevate Labs

Student Name: Amar H M

Internship Program: AI & ML Internship

Task 10: KNN – Handwritten Digit Classification

Date: 19/1/2026

Tool Used: Python (Scikit-learn) using VS Code (Jupyter Notebook)

Dataset Used: Scikit-learn Digits Dataset

1. Introduction

This report presents the classification modeling process performed as part of Task-10 of the AI & ML Internship conducted by Elevate Labs. The main objective of this task is to implement the K-Nearest Neighbors (KNN) algorithm to classify handwritten digits and evaluate its performance.

2. Dataset Overview

The Digits dataset is a built-in dataset provided by Scikit-learn. It consists of grayscale images of handwritten digits from 0 to 9. Each image is represented as an 8×8 pixel grid and flattened into numerical features.

- **Dataset Type:** Image-based structured data
 - **Total Samples:** 1797
 - **Target Classes:** Digits (0–9)
-

3. Data Visualization

Sample images from the dataset were visualized to understand the structure of the data and confirm correct labels. This helps in verifying the quality and diversity of the dataset.

4. Train–Test Split

The dataset was divided into training and testing sets using an 80:20 ratio. The training set was used to train the model, while the testing set was used for evaluation.

5. Feature Scaling

StandardScaler was applied to normalize feature values. Since KNN is a distance-based algorithm, scaling ensures that all features contribute equally during distance computation.

6. Model Training

A KNN classifier was trained using $K = 3$. The model predicts the class of a new sample based on the majority class among its nearest neighbors.

7. Hyperparameter Tuning

Multiple K values (3, 5, 7, 9) were tested to observe their effect on accuracy. This helps in selecting an optimal K value for better generalization.

8. Model Evaluation

The model was evaluated using accuracy score and confusion matrix. These metrics provide insight into how well the model classifies each digit.

9. Confusion Matrix Analysis

The confusion matrix shows the number of correct and incorrect predictions for each digit. It helps in identifying which digits are more frequently misclassified.

10. Interpretation of Results

The KNN model achieved high accuracy in classifying handwritten digits. The results indicate that KNN performs well on small, well-structured image datasets. The accuracy vs K plot shows how increasing K affects model performance.

11. Conclusion

The KNN algorithm was successfully implemented for handwritten digit classification. The dataset was visualized, scaled, split, and evaluated using standard metrics. This task provided practical understanding of distance-based learning and the impact of hyperparameters on model performance.
