# K-Nearest Neighbors (KNN) Classification Report for MNIST Dataset Name: Khizar Hayat Reg. No: M24F0062DS011 Name: Umer Sajid Reg. No: M24F0019DS005

Group: 07

## Abstract

This report presents the implementation and evaluation of the K-Nearest Neighbors (KNN) algorithm on the MNIST dataset. The study aims to identify the optimal distance metric and the value of K by comparing the model's performance using evaluation metrics such as accuracy and confusion matrices.

## Introduction

The K-Nearest Neighbors (KNN) algorithm is a fundamental supervised machine learning technique. This report evaluates KNN on the MNIST dataset, which comprises handwritten digit images. The objective is to classify these digits accurately using different distance metrics (Euclidean, Manhattan, Chebyshev) and varying values of K (from 1 to 10).

## Dataset

The MNIST dataset contains 70,000 grayscale images of handwritten digits, divided into 60,000 training samples and 10,000 testing samples. Each image is of size 28x28 pixels, representing digits from 0 to 9. The dataset is widely used for benchmarking image classification algorithms.

## Model

The KNN model was implemented using the Scikit-learn library. The algorithm computes distances between data points and their neighbors to classify a given input based on the majority class of its nearest neighbors. Different distance metrics (Euclidean, Manhattan, Chebyshev) and values of K (ranging from 1 to 10) were tested.

## Results

### Evaluation Metrics

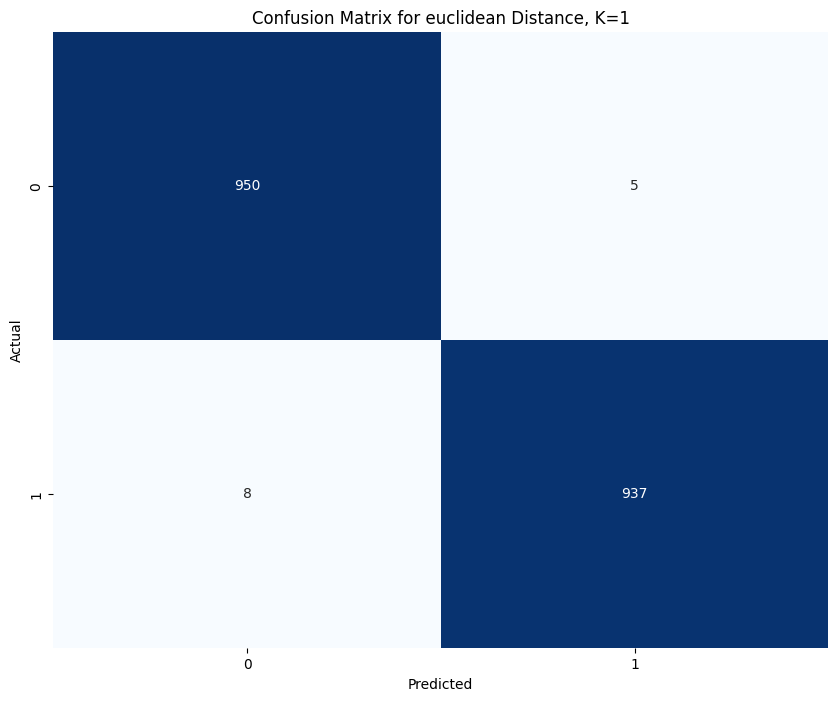
The evaluation of the KNN model is based on accuracy and confusion matrices. The accuracy metric measures the percentage of correctly classified samples, while confusion matrices provide insights into the model's performance for each class.

|  |  |  |
| --- | --- | --- |
| K Value | Distance Metric | Accuracy |
| 1 | Euclidean | 0.972 |
| 5 | Manhattan | 0.965 |
| 10 | Chebyshev | 0.8106428571428571 |

A comparison of accuracy for different metrics and K values is summarized in the table above.

### Confusion Matrix

Below is the confusion matrix for the Euclidean distance metric with K=1, which achieved one of the highest accuracies.



## Conclusion

Based on the evaluation, the Euclidean distance metric with K=1 provided the highest accuracy for the MNIST dataset. The results demonstrate that the choice of K and distance metric significantly impacts the performance of the KNN algorithm.