Summary

This project implemented handwritten digit recognition for banking applications using the MNIST dataset. Three classifiers—Gaussian Naïve Bayes, Non-Naïve Bayes, and KNN—were trained and evaluated to identify digits from deposit slips. Data preprocessing included feature scaling and imbalance handling to improve accuracy. The models were compared based on accuracy and computational efficiency.

Programming Language: Python

Libraries: NumPy, Pandas, Matplotlib, seaborn

Machine Learning: Custom implementation of Bayes and KNN classifiers

Evaluation Metrics: Accuracy, Confusion Matrix

MNIST dataset:

The MNIST dataset, consisting of 70,000 handwritten digits (0-9), was used to train classification models. Three approaches were implemented:

Gaussian Naïve Bayes (GNB): Fast and efficient but assumes feature independence. GNB improved from 59% to 77% accuracy with feature scaling.

Non-Naïve Bayes (Multivariate Gaussian): Accounts for pixel correlations for better accuracy. Non-Naïve Bayes reached 95% accuracy after parameter adjustments.

K-Nearest Neighbors (KNN): Provides the highest accuracy but is computationally expensive. KNN achieved the highest accuracy of 97% (Referenced), making it the best but slowest model.

The results showed that KNN performed the best in accuracy, while Naïve Bayes provided a balance between speed and accuracy. The study highlights the potential of machine learning in financial automation, reducing fraud, and improving banking efficiency.