Handwritten Recognition in Banking : MNIST Implementation

Prepared By: Vaishali Patelia

Content

> Introduction

> Problem

> MINST Dataset

> Solutions - Machine Learning Approaches

> Implementation

> Conclusion

Handwritten Filled Deposit Slip

Source: supermoney

Problem

Handwritten Deposit Slip

Manual check and deposit slip processing is slow and error-prone.

Document verification is inefficient.

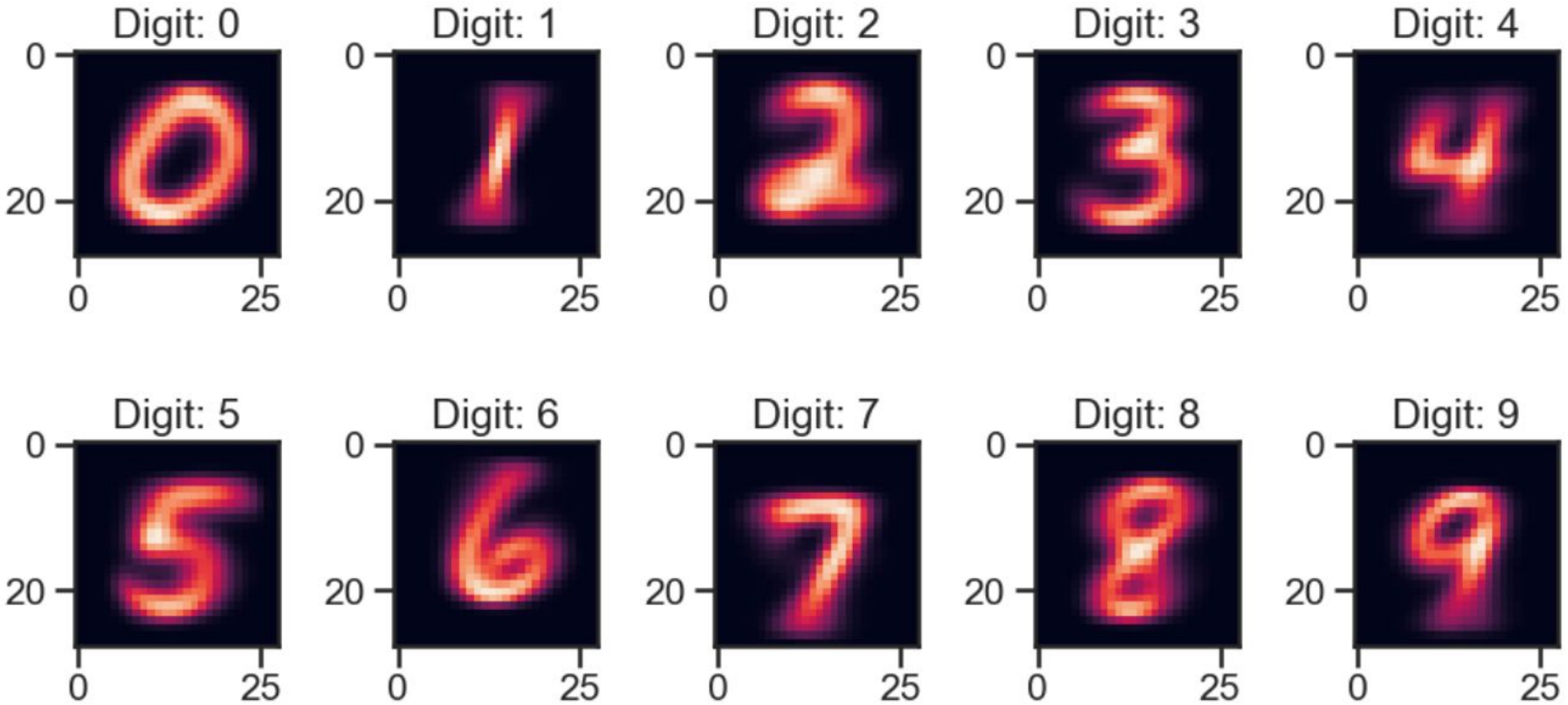
MNIST Dataset Overview

- Number of Instances & Attributes: 70,000 images & 784 (28x28 pixels)
- Target: Column represents the digit (0-9) corresponding to the handwritten image

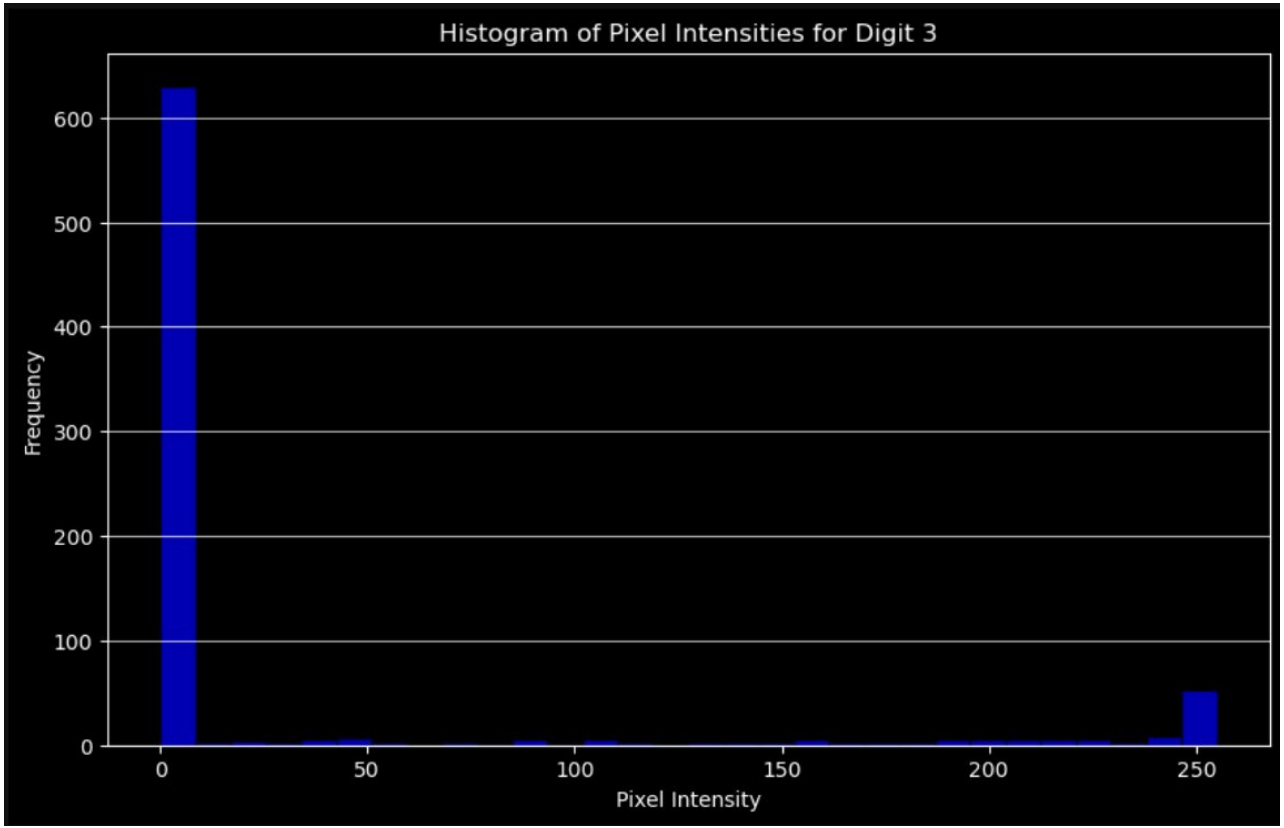
The dataset is divided into two main subsets:

- Training Set: Consists of 60,000 images along with their labels.
- Test Set: Contains 10,000 images with their corresponding labels.

Digits from the MNIST dataset

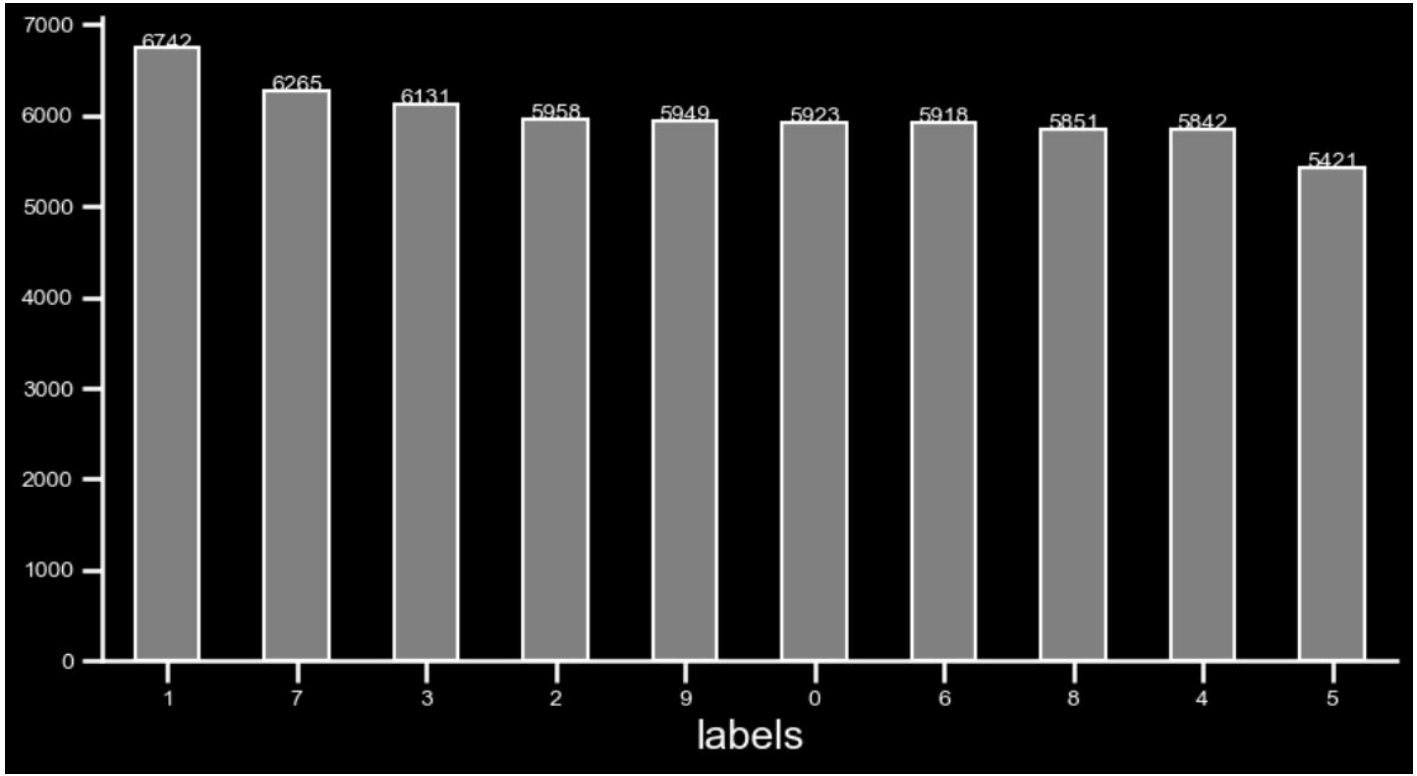


Visualization of Mean Pixel Intensity Across MNIST Dataset:
This image represents the average pixel intensity for each digit (0-9)



Histogram of Pixel Intensities for Digit 3

Checking Target Imbalance



Solutions

Machine Learning Approaches

Gaussian Naïve Bayes (GNB)

Non-Naïve Bayes (Multivariate Gaussian)

K-Nearest Neighbors (KNN)

Gaussian Naïve Bayes (GNB)

Used Scaling to Improve Accuracy

Accuracy	Before Scaling	After Scaling - X/255
Train Dataset	0.593	0.768
Test Dataset	0.587	0.774

Non-Naïve Bayes (Multivariate Gaussian)

Adjust epsilon to Improve Accuracy

Accuracy	Before Adjust Parameter	After Adjust Parameter
Train Dataset	0.783	0.9549
Test Dataset	0.751	0.9542

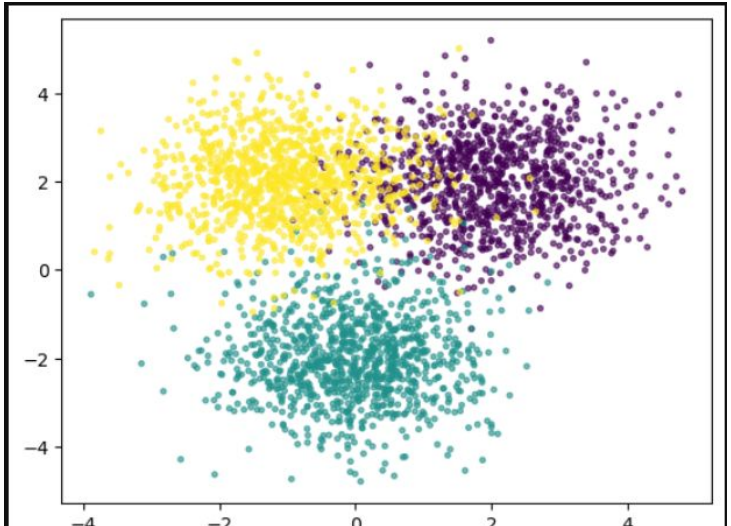
K-Nearest Neighbors (KNN)

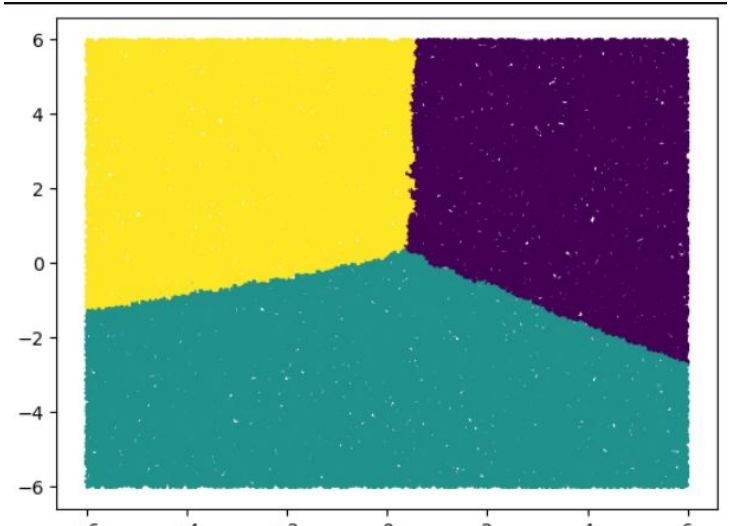
Finds closest k points using distance. K = 20 / 200

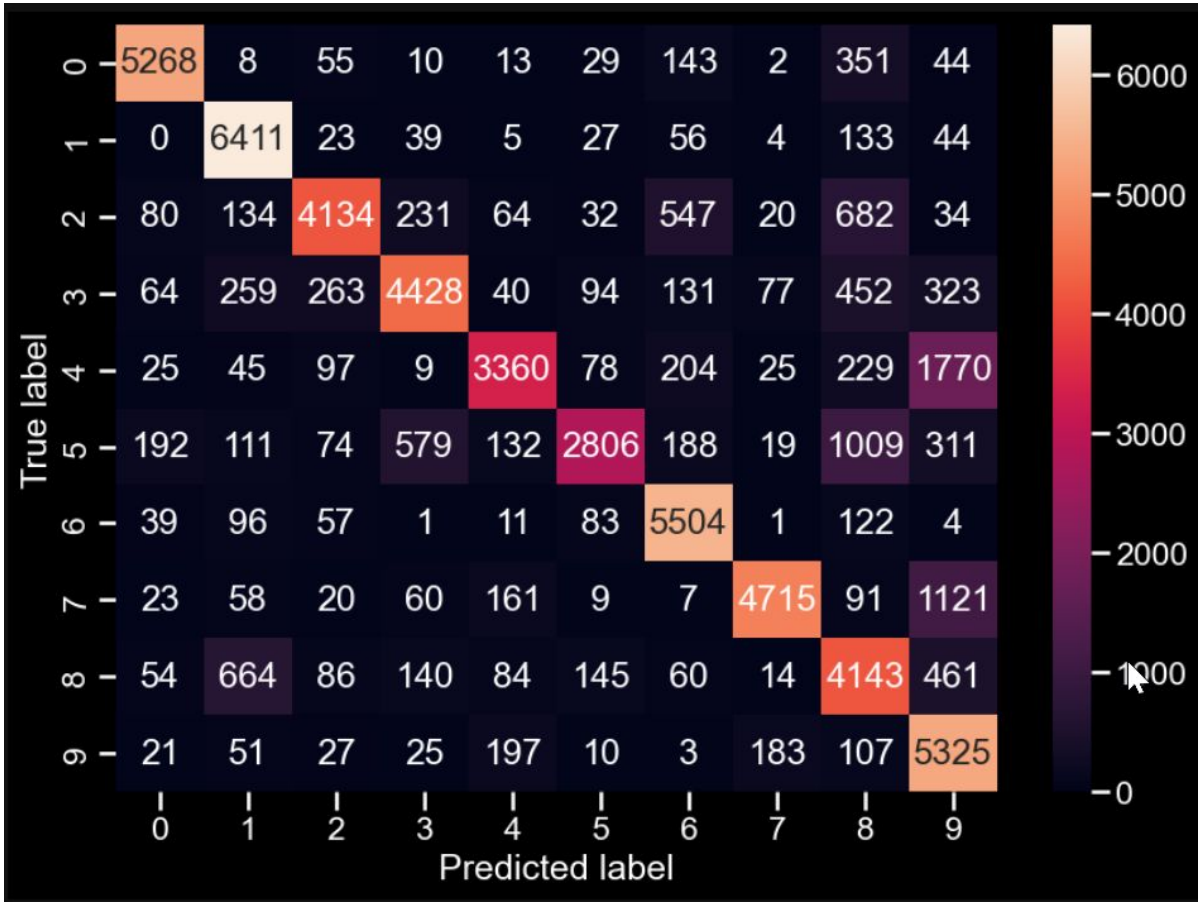
Extracts a subset of the training dataset

Accuracy of train dataset - 0.902 test dataset - 0.91

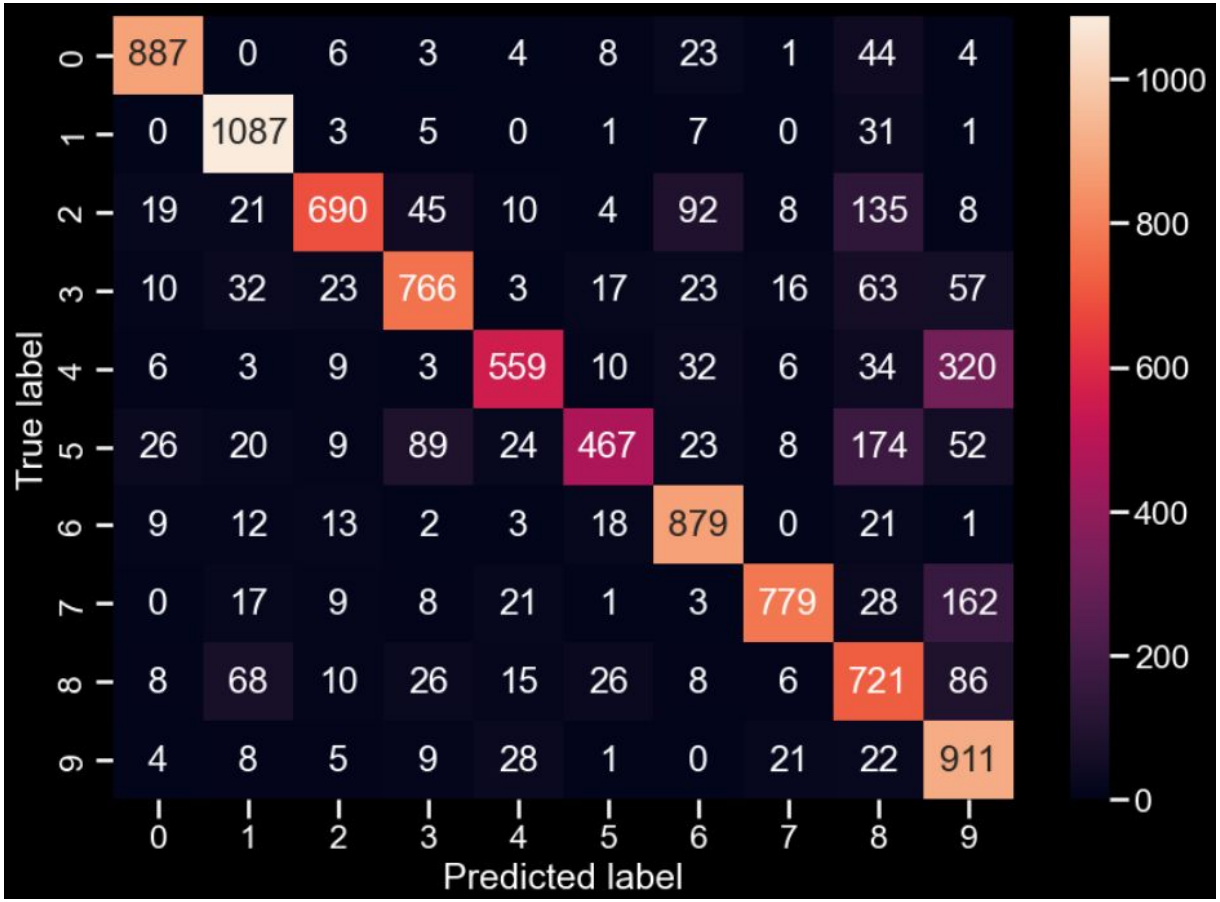
Using Full Dataset (70,000 samples): ~97% (Referenced)







Confusion Matrix of Train Dataset



Confusion Matrix of Test Dataset

Model Training & Results

Algorithm	Accuracy	Pros	Cons
Gaussian Naïve Bayes	77.4% (after scaling)	Fast, low memory usage, improved accuracy with scaling	Assumes feature independence
Non-Naïve Bayes	95.42% (after adjust parameter)	Captures pixel dependencies	Complex computation
KNN	97% (Referenced)	High accuracy	Slow inference time

Implementation

Mobile Check Deposit using Handwriting Recognition

Banking Apps (Chase, Wells Fargo, PayPal, Google Pay, Apple Pay)

Source: callawaybank

Mobile Check Deposit App

How it Works:

User takes a picture of a check.

Machine learning model detects & extracts handwritten account numbers and amounts.

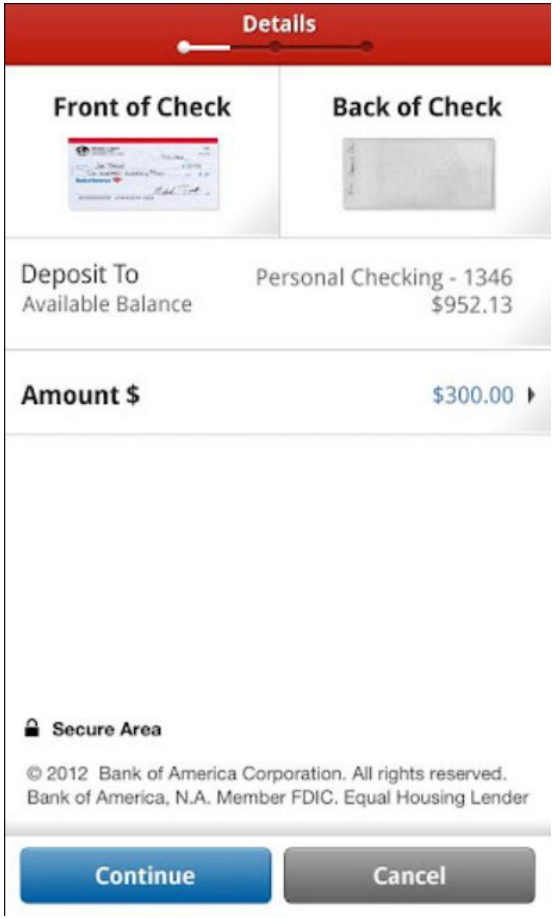
Verifies authenticity and detects fraud.

Deposits funds into the account automatically.

Benefits:

Reduces fraud & manual verification.

Faster processing & improved user experience.



Source: androidauthority

Conclusion

Handwritten recognition significantly enhances the speed and accuracy of processing checks and receipts.

GNB improved from 59% to 77% accuracy with feature scaling.

Non-Naïve Bayes reached 95% accuracy after parameter adjustments.

KNN achieved the highest accuracy of 97%, making it the best but slowest model.


Thank you!