Exam Paper Evaluation Using the MNIST Dataset

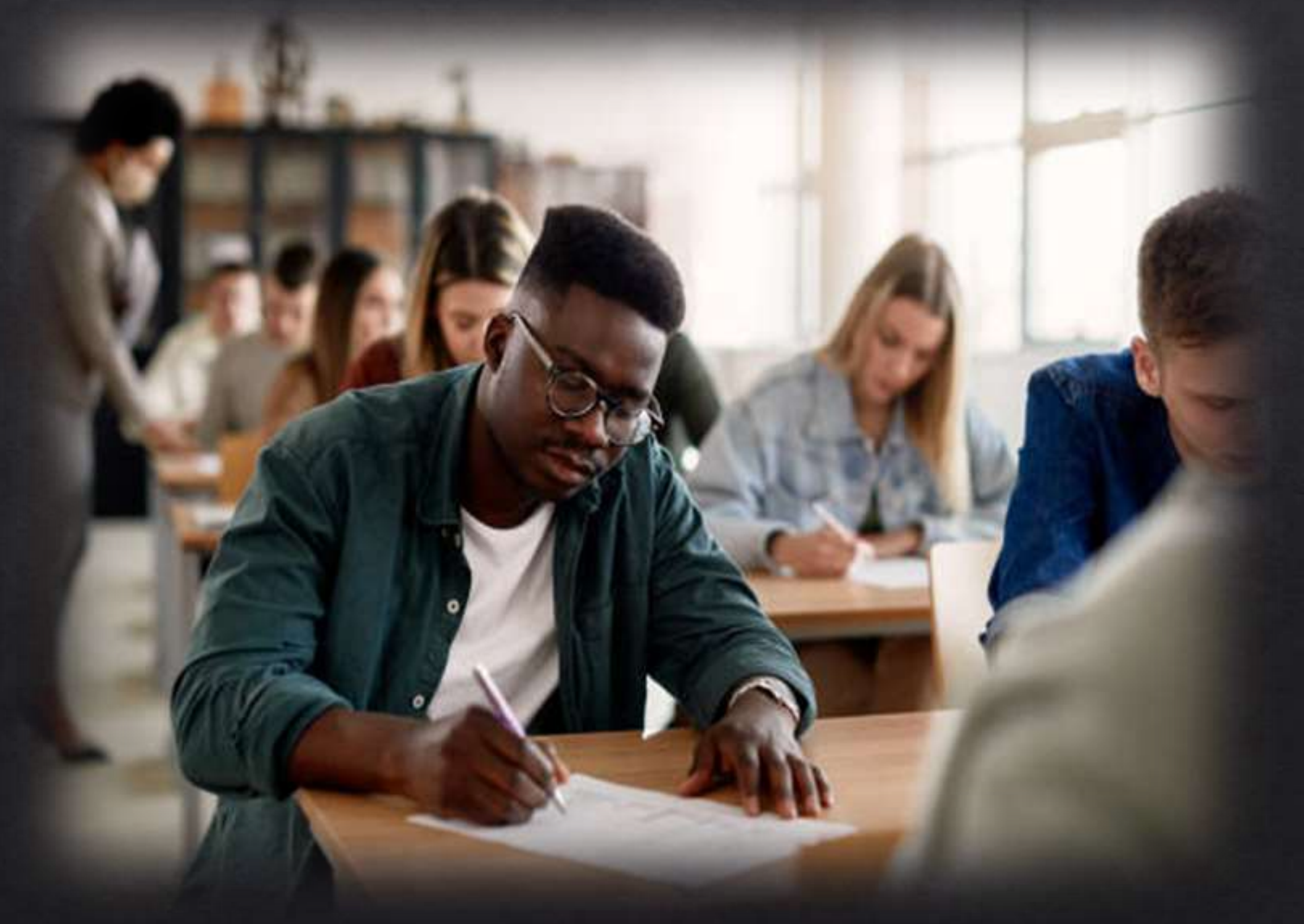
Automatic Recognition and Grading of Handwritten Digits Using Bayes Classifiers

Presented by: Pascal Opara

Content Overview

Use Case Summary

This project simulates how machine learning can help in automatic exam paper grading. The goal is to classify handwritten digits, similar to how students write numeric answers or select multiple-choice options. Using the MNIST dataset, two types of Bayes classifiers were applied to detect and recognize digits. This helps human effort, minimize grading errors, and speed up the evaluation process.



Dataset Description

- MNIST Dataset Used
- Training Samples: 60,000
- Testing Samples: 10,000
- Data Format: 28x28 grayscale images of digits (flattened into 784 pixel features)
- Pixel Range: 0 to 255
- Label Range: 0 to 9 (representing digits)

Processes & Justification

Step 1

Loaded MNIST_train.csv and MNIST_test.csv files and cleaned them by removing unnamed or duplicate columns

Step 2

Preprocessing:

- Raw Data – No scaling
- Normalized Data – pixel values were divided by 255 to scale them from 0-255 to 0-1

We normalized because Naïve Bayes assumes normal distribution for pixel values. It makes them suitable for this assumprion and improves accuracy

Step 3

Model Implementation:

Naïve Bayes Classifier (GausNB): Assumes independence between features and uses mean and variance.
Non-Naive Bayes Classifier(guass_non_nb):
Uses full covariance matrices for each class and does not assume independence

Step 4

Model Training

Used the .fit() method on both models to learn the necessary parameters (mean, variance, or full covariance) from the training data.

Step 5

Testing & Evaluation:
Used .predict() on the test dataset to make predictions, and evaluated performance by calculating accuracy against true labels

Modeling Techniques

Non_Naive Bayes(guass_non_nb)

More complex - uses covariance matrices for each class. It does not assume feature independence but is more prone to overfitting

Naive Bayes(GausNB)

Fast, probabilistic model that works well on normalised data. Assumes features are independent and uses per-class mean and variance

Both models were applied to Raw Pixel Data & Normalized Pixel Data

Final Observations

Insights:

- Naïve Bayes performed poorly on raw data but improved significantly when normalized.
- With normalization, it achieved a strong and balanced performance on both training and test sets.
- Non-Naïve Bayes performed better because it considered covariance between pixels.
- Non-Naïve Bayes was more consistent and generalizable across data formats

Model Variant	Train Accuracy	Test Accuracy
Naïve Bayes (raw)	59.38%	-
Naive Bayes (normalized)	76.82%	77.46%
Non_Naive Bayes (raw)	78.56%	75.32%
Non-Naive Bayes (normalized)	93.06%	91.08%

Obstacles Faced

- Raw Data Limitations

Naïve Bayes did not work well on raw data because the pixel values did not match its assumptions.

- Training Time

Training time was long when working with 60,000 rows

- Data Cleaning

Extra or unnamed columns were removed before model training.

- Limited visuals

Only prediction outputs were available. There were not enough visuals to help explain model behaviour clearly.

Final Recommendations

- Normalize pixel data before using it in models like Naive Bayes.

• Start with simple models — they are faster and often work better with fewer resources.

• Don’t rely only on training accuracy — always test with new data.

• Avoid overfitting by checking if a complex model performs poorly on test data.

• Add simple visuals or charts that compare predictions — they help explain your results clearly.

Conclusion

This project showed how a machine learning model can automate grading of digit-based exam responses. We successfully implemented digit recognition using Naïve and Non-Naïve Bayes classifiers on the MNIST dataset. Normalization was key to achieving high accuracy, with the best result being 91.08% using Non_naive Bayes. This proves that automatic exam grading using handwritten digit recognition is practical and reliable.

Q&A

Any Questions?