

Handwritten Digit Recognition using KNN and Naive Bayes

Muddasar Razzaq
Employee of Information Tech Consultants
Ulster University

Objective & Dataset Overview

■ Objective:

- Recognize handwritten digits (0–9) using the MNIST dataset.
- Compare K-Nearest Neighbors (KNN) and Gaussian Naive Bayes (GNB) models.

■ Dataset:

- 60,000 training and 10,000 testing images.
- Each image is 28x28 pixels (784 features).
- Labels represent digits from 0–9.

Data Preprocessing

■■ Steps:

1. Removed unnecessary columns (e.g., 'Unnamed: 0').
2. Normalized pixel values from [0, 255] → [0, 1].
3. Used provided train/test split.
4. Applied PCA for dimensionality reduction in KNN.

Model Comparison: KNN vs Naive Bayes

■ Gaussian Naive Bayes (GNB):

- Accuracy: ~60%
- Fast and simple, assumes feature independence.
- Good for quick baselines.

■ K-Nearest Neighbors (KNN):

- Accuracy: ~80%
- Distance-based, non-parametric method.
- More accurate but slower on large datasets.

Conclusion

■ Key Takeaways:

- Both models classified digits successfully.
- KNN achieved higher accuracy (~80%) vs Naive Bayes (~60%).
- Naive Bayes is faster and simpler to implement.
- Demonstrates trade-off between accuracy and computational cost.

- Project shows effective preprocessing and model comparison in supervised learning.