# HW1 Analysis Report

**Introduction**

This report analyzes the dataset used in HW1, focusing on machine learning techniques applied to digit classification. The dataset underwent preprocessing using Principal Component Analysis (PCA) and Standard Scaling before evaluating models such as K-Nearest Neighbors (KNN) and Naïve Bayes.

**Dataset Overview**

* **Training Set:** 7,291 samples with 256 features each
* **Test Set:** 2,007 samples with 256 features each
* **Classes:** 10 unique digit labels (0-9)

**Data Preprocessing**

1. **Dimensionality Reduction:** PCA was applied to reduce features to 50 principal components.
2. **Feature Scaling:** StandardScaler was used to normalize data.

**Model Evaluation**

**K-Nearest Neighbors (KNN)**

* **Cross-validation Accuracy:** 97.3%
* **Test Accuracy:** 94.86% (best with k=3)
* **Precision Score:** 0.9484
* **Recall Score:** 0.9435
* **Confusion Matrix:** Provided for k-values 1-10, showing classification performance.
* **Observations:**
  + Best accuracy with k=3
  + Higher k-values slightly reduced accuracy

A graph with blue and orange lines

AI-generated content may be incorrect.

**Naïve Bayes Classifier**

* **Train Accuracy:** 90.09%
* **Test Accuracy:** 86.70%
* **Precision Score:** 0.8603
* **Recall Score:** 0.8568
* **Confusion Matrix:** Displayed for performance insights.
* **Observations:**
  + Lower accuracy compared to KNN
  + Struggled with misclassification of certain digits

**Model Comparison**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Train Accuracy** | **Test Accuracy** | **Precision** | **Recall** |
| KNN (k=3) | 97.3% | 94.86% | 0.9484 | 0.9435 |
| Naïve Bayes | 90.09% | 86.70% | 0.8603 | 0.8568 |

**Conclusion**

KNN (with k=3) significantly outperformed Naïve Bayes in terms of accuracy, precision, and recall. While Naïve Bayes is computationally efficient, its assumptions about feature independence led to weaker performance.

This analysis highlights the effectiveness of KNN for digit classification while suggesting areas for further refinement.

*Team 17:*

* *Arshad Jafri*
* *Raghu Vamsi*
* *Sonu Soni*