

# Project Report: Feature Selection and Classification using Naive Bayes

## Objective

The main objective of this project was to perform feature selection and classification on the MNIST dataset using a Naive Bayes classifier. The project aimed to identify the optimal subset of features that would yield the highest classification accuracy.

## Methodology

The project followed these steps:

1. **Data Loading:** The MNIST dataset was loaded using the `tensorflow.keras.datasets` module.
2. **Data Preprocessing:** The images in the dataset were reshaped from 28x28 pixels to a 1D array of 784 features. Features with zero variance were removed using `VarianceThreshold` from `sklearn.feature_selection`.
3. **Feature Selection:** Forward feature selection was performed to select the best features for the Naive Bayes classifier. For each feature in the dataset, a Naive Bayes classifier was trained using that feature along with the features already selected, and its performance was evaluated on the validation set. The feature that gave the best performance was selected.
4. **Model Training and Evaluation:** A Naive Bayes classifier was trained using the selected features and its performance was evaluated on the test set.

## Results

The Naive Bayes classifier was able to achieve a final accuracy on the test set using the selected features. The accuracy of the model increased as more features were added, indicating that the forward feature selection process was successful in identifying relevant features.

## Conclusion

This project demonstrated the effectiveness of forward feature selection in improving the performance of a Naive Bayes classifier on the MNIST dataset. The visualizations provided valuable insights into the model's performance and the importance of the selected features. The techniques used in this project could be applied to other datasets and models to improve their performance.