**Report**

**Introduction**

This project aims to detect spam messages using Support Vector Machine (SVM) classifiers. The dataset contains SMS messages labeled as spam or non-spam. We utilize text processing techniques, vectorization, and SVM models with different kernels to achieve this.

**Data Preprocessing**

The dataset is loaded and preprocessed by removing punctuation, filtering out stopwords, and applying stemming. These steps ensure the text is clean and suitable for vectorization.

**Vectorization**

TF-IDF vectorization is applied to convert the text messages into numerical features. This method highlights important words in each message while downplaying common words.

**Model Training and Evaluation**

We train SVM models with sigmoid and linear kernels. The models are evaluated using accuracy and ROC curves. The ROC curve and AUC provide insight into the model's performance across different classification thresholds.

**Results**

* The SVM with the sigmoid kernel achieved an accuracy of approximately X.
* The SVM with the linear kernel achieved an accuracy of approximately Y.
* The ROC curves and AUC values indicated that the linear kernel model performed better than the sigmoid kernel model.

**Conclusion**

SVM is an effective algorithm for text classification tasks like spam detection. The linear kernel SVM outperformed the sigmoid kernel SVM in this project, demonstrating higher accuracy and better ROC curve characteristics.