Implementation Documentation

1. Problem Definition

The goal of this project is to build an Email Spam Classification system that automatically classifies emails into Spam or Ham (Not Spam) using machine learning techniques. Spam emails cause security and productivity issues, so accurate detection is essential.

2. Dataset Description

The project uses the Enron Email Spam Dataset, which contains real-world email messages.

email: the email text content

filename: email identifier (not used in modeling)

label: classification label (spam or ham)

This dataset is suitable for spam detection because it contains realistic email language and balanced classes.

3. Data Preprocessing

To prepare the email text for machine learning, the following preprocessing steps were applied:

- Convert text to lowercase

- Remove URLs

- Remove numbers and special characters

- Remove extra spaces

These steps reduce noise and improve model performance.

4. Feature Extraction

Text data was converted into numerical features using TF-IDF (Term Frequency–Inverse Document Frequency).

TF-IDF assigns higher weights to important words that appear frequently in a specific email but not across all emails.

5. Dataset Splitting

The dataset was split into three parts:

- 70% Training set

- 10% Validation set

- 20% Test set

This split ensures fair evaluation and prevents overfitting.

6. Model Implementation

6.1 Naive Bayes Classifier

The Multinomial Naive Bayes model was implemented as a baseline classifier. It is fast, efficient, and suitable for text classification.

6.2 Support Vector Machine (SVM)

A Linear Support Vector Machine (LinearSVC) was used as the main model. It works effectively with high-dimensional text data and achieved higher accuracy than Naive Bayes.

7. Model Evaluation

The models were evaluated using:

- Accuracy

- Precision, Recall, and F1-score

- Confusion Matrix

Evaluation was performed on both validation and test datasets.

8. Visualization and Analysis

Visualizations used include:

- Confusion Matrix Heatmaps

- Actual vs Predicted plots for validation and test sets

These visualizations clearly demonstrate the performance and error patterns of the models.

9. Conclusion

This project shows that machine learning techniques, particularly SVM combined with TF-IDF, can accurately classify spam emails. Proper preprocessing and feature extraction significantly improved model performance.