PA LAB ASSIGNMENT:
SUBMITTED TO ACHALA SHAKYA MA'AM
DONE BY: SPARSH BARSE
ANANYA AGRAWAL

Abstract/Introduction

With the rapid growth of digital communication, spam emails and messages have become a prevalent nuisance. A Ham and Spam Classifier uses machine learning to differentiate between legitimate (ham) and unwanted (spam) messages. This study explores a robust classification model to improve email and message filtering, ensuring better security and user experience.

Problem Statement

Unwanted spam emails not only clutter inboxes but also pose cybersecurity risks by spreading malware or phishing links. Traditional rule-based filters often fail to adapt to evolving spam techniques, leading to inefficiencies. Therefore, a need arises for a reliable machine learning-based system to classify messages as ham or spam with high accuracy.

Predictions

The model is expected to:

- Achieve over 90% accuracy in classifying messages.
- Handle previously unseen spam patterns effectively.
- Minimize false positives (ham classified as spam).
- Maintain computational efficiency for real-time applications.

Challenges Faced

- Data Imbalance: Datasets often contain significantly more ham messages than spam.
- Evolving Spam Tactics: Spammers constantly adapt their techniques to bypass filters.
- Feature Selection: Finding the most relevant features from the dataset, such as keywords or sender behavior, is challenging.
- Misclassification Penalties: False positives may lead to the loss of important emails, impacting user trust.

Future Scope

- Integration with advanced Natural Language Processing (NLP) techniques like transformers for improved accuracy.

- Application to multi-lingual spam classification.
- Enhancements for detecting image or multimedia-based spam.
- Real-time implementation for SMS, chat applications, and other digital platforms.

Conclusion

The Ham and Spam Classifier demonstrates the potential of machine learning in addressing spam-related issues effectively. With further refinements, it can offer a seamless and secure user experience by automating the filtration of spam messages.

Visualizations and Insights

- Distribution of ham vs. spam messages in the dataset.
- Word cloud of common keywords in spam messages.
- Confusion matrix showing classification performance.
- Accuracy, precision, recall, and F1 score trends across different model iterations.

Dataset Description

- Source: Public datasets like SMS Spam Collection or Enron Email Dataset.
- Structure: Text of the message, label (ham/spam).
- Preprocessing: Cleaning, tokenization, and stopword removal.

Methodology

1. Data Collection: Import and analyze the dataset.
2. Data Preprocessing: Cleaning, text normalization, and tokenization.
3. Feature Extraction: Using techniques like TF-IDF or word embeddings.
4. Model Selection: Testing algorithms like Naïve Bayes, SVM, or deep learning models.
5. Evaluation: Metrics like accuracy, precision, recall, and F1 score.
6. Optimization: Fine-tuning hyperparameters for optimal performance.

Objectives

- Develop a classifier capable of distinguishing between ham and spam messages.
- Optimize the model for real-world implementation.
- Provide insights into common spam patterns and their evolution.