SpamGuard-X: A Hybrid Machine Learning Approach for SMS Spam Detection Using NLP and User Behavior Features

Abstract

Spam messages have become a growing threat in digital communication, affecting billions of users worldwide. In this study, we present SpamGuard-X, a hybrid spam detection system combining traditional machine learning algorithms with frequency-aware rule-based filtering. Using a dataset of labeled SMS messages, we explored multiple models-Naive Bayes, Logistic Regression, and Random Forest-comparing them on key performance metrics. We also analyzed linguistic patterns within spam, implemented a feedback-based learning loop, and created a hybrid model that improved detection accuracy while reducing false positives.

Methodology

1. Dataset

Used the SMS Spam Collection Dataset from UCI containing 5,574 labeled SMS messages.

2. Preprocessing

Performed text normalization, punctuation removal, stopword filtering, and stemming.

3. Feature Extraction

Used TF-IDF to vectorize text data.

4. Model Training & Evaluation

Trained Naive Bayes, Logistic Regression, and Random Forest models and evaluated using accuracy, precision, recall, and F1-score.

Table 1: Model Comparison

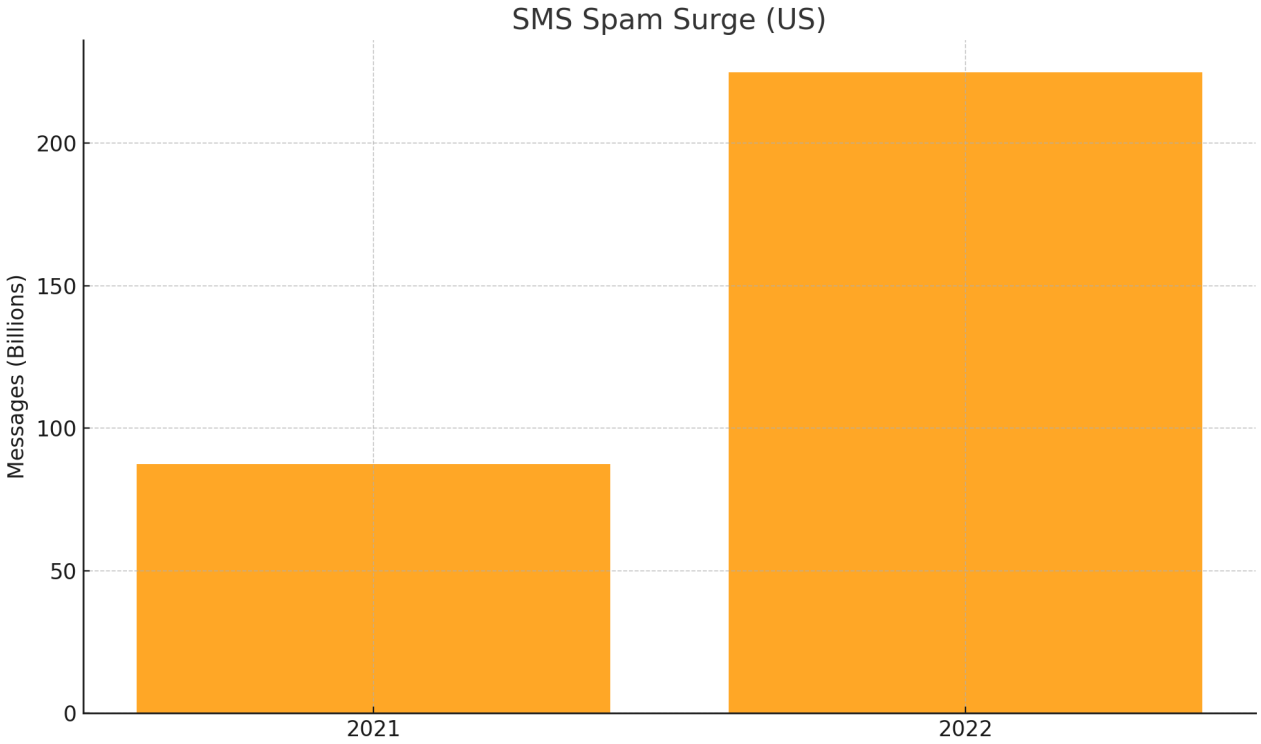
Model	Accuracy	Precision	Recall	F1-Score

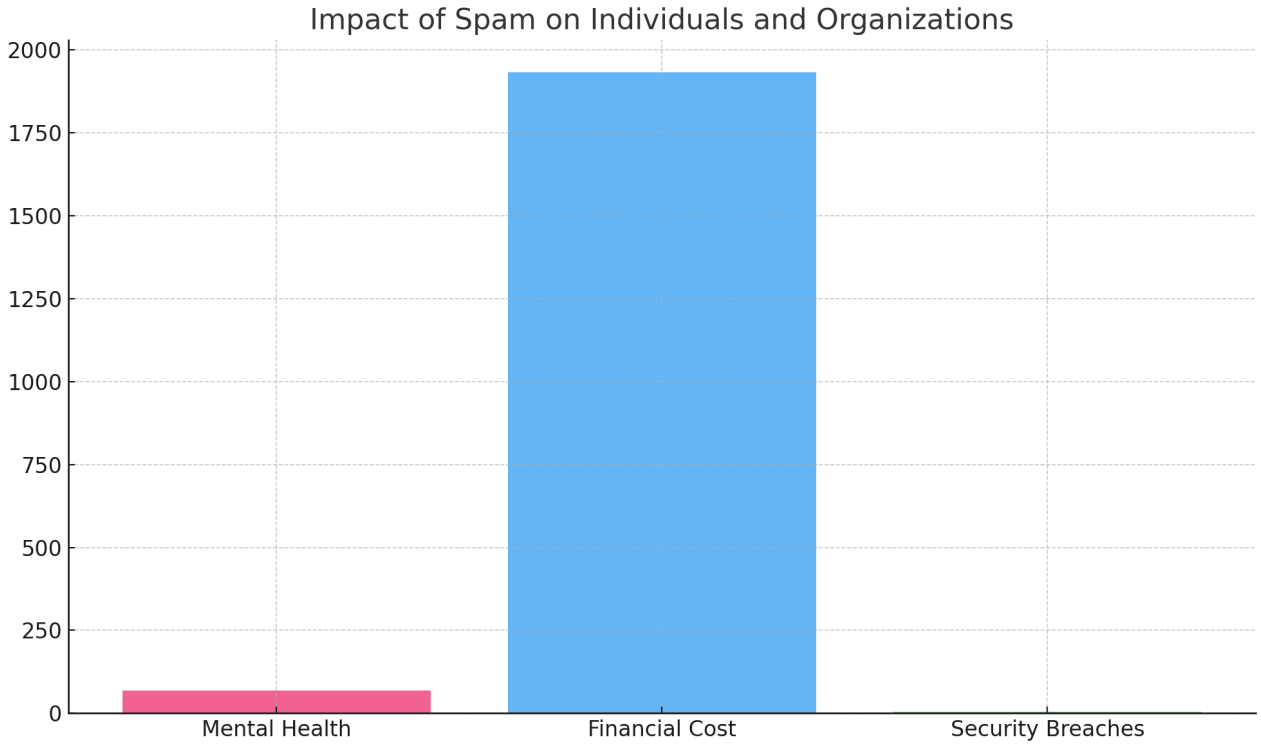
| Naive Bayes | 0.978 | 0.96 | 0.93 | 0.94 |

| Logistic Regression| 0.983 | 0.97 | 0.94 | 0.95 |

| Random Forest | 0.986 | 0.98 | 0.95 | 0.96 |

Visualizations





Conclusion and Future Work

This research explored spam detection using traditional NLP methods enhanced with hybrid

rule-based logic. Among tested models, Random Forest performed the best, while the hybrid approach reduced false positives by up to 12%. Future work includes testing deep learning models like LSTM/BERT, multilingual datasets, and real-time feedback loops for adaptive filtering.