

Hybrid SMS and Email Fraud Detection using Machine Learning (Naïve Bayes + SVM)

Ms. Ananya Singh¹

B. Tech (Information Technology) NIET, Greater Noida

Email: 0221ite260@niet.co.in

Ms. Mini Jain²

Assistant Professor (IT Department) NIET, Greater Noida

Email: mini.jain@niet.co.in

Abstract

Owing to its alarming volume, spam and fraud messages sent through SMS and e-mail have been among the most predominant threats faced to cyber security in recent years. Bad guys send out spam containing malicious content for phishing, financial fraud, identity theft, and malware distribution. Traditional spam filters and rule-based systems don't work well at adapting to changing spam strategies.

In this paper, the author puts forward a hybrid detection model, which integrates Naïve Bayes and Support Vector Machines (SVM) together with Term Frequency- Inverse Document Frequency (TF-IDF) vectorization. Naïve Bayes provides fast classification and SVM provides high dimensional feature space separation. The proposed model has better accuracy, lower false positive rate than single algorithm models.

The results of the experiments performed on the SMS Spam Collection and Enron Email Dataset give an accuracy value of

~97% with the hybrid approach, significantly improving precision and recall. The paper presents the preprocessing pipelines, evaluate metrics, real world applications and challenges for large scale deployment.

Keywords: EFL classrooms, Spam Detection, Email Fraud, Naïve Bayes, SVM, Machine Learning, Cybersecurity, Hybrid Models

1. Introduction

The rapid development of digital technology has changed the way people communicate, conduct business and share information. Email and Short Message Service (SMS) are used because they are easily accessible, inexpensive, and fast. This omnipresence has also made them the subject of messages distribution and frauds.

With more than 300 billion emails being sent every day, estimates put around 45- 50% of those at spam. Mobile networks note a gradually growing number of fraudulent SMS messages every year, particularly in areas where mobile-based financial transactions are popular. These statistics make a cry for help for advanced detection systems.

Spam and phishing messages have now become sophisticated and intelligent. The most common types are phishing (which involves tricking people into disclosing their passwords or banking information), smishing (sending spam links via SMS to install malware) and Business Email Compromise (BEC). The FBI's IC3 reports

that losses from email fraud from 2016 to 2023 cost over \$26 billion.

The conventional spam filters using rules and keyword blacklists are inflexible and are easily defeated by spammers' ability to tweak content or sprinkle in keywords. Machine learning (ML) has proved to be a

powerful solution as it can learn from large volumes of data, generalize to unseen data, and improve via retraining.

Naïve Bayes is well known for its simplicity and efficiency. SVM is known to be robust in high-dimensional feature space which is typically used in NLP following TF-IDF vectorization. Hybrid can be used to use Naïve Bayes as a fast first-pass classifier followed by the SVM to resolve the ambiguous decisions to minimize false positives and negatives.

In this paper, a hybrid spam detection system is designed, implemented and evaluated using Naïve Bayes and SVM, which is more accurate than both algorithms when evaluated in terms of accuracy, precision, recall and F1-Score.

2. Literature Review

Spam detection is one of the most prominent areas of research for the past 20 years. Early spam filters were rule-based, with handcrafted keyword lists. These were easy to implement, but were not very adaptable in the face of the spammers' evolution in tactics.

Naïve Bayes classifier is one of the first popular ML algorithms used in text classification. Dharani et al. (2023) showed that accuracy of Naïve Bayes on SMS dataset is about 95%, and it is suitable to be used on lightweight devices.

An area that SVMs excelled at and became popular was their performance with high-dimensional feature spaces. Vasudevan et al. (2024) proposed SVM based email Fraud Detection technique with 97.5% accuracy. Using ensemble methods like Random Forest and AdaBoost is more accurate but requires more computational complexity, limiting their use for real time applications.

The deep learning methods (CNNs, RNNs) are trained without any need for human interaction, but they need huge quantities of training data, high computation and are not always interpretable. Literature survey shows that hybrid algorithms based on different complementary algorithms are superior to single algorithm-based models, which is the motivation of this study.

3. Problem Statement

The widespread use of digital communication systems like email, SMS and social media has also improved the proliferation of spam messages, phishing, malware, identity theft and financial fraud.

The traditional, keyword-based and heuristic filters are easy and ineffective against new spam techniques like text obfuscation and contextual camouflage.

The traditional machine learning techniques such as Naive Bayes, Support Vector Machines (SVM), Decision Trees and Ensemble methods have been found to be effective in spam detection, but they

have some drawbacks: low adaptability, low recognition rate, inability to handle large amount of data, and inability to

recognize imbalanced datasets. Naïve Bayes is fast to classify but assumes

independence while SVM has strong discriminative power at the expense of higher computation cost. Hence, it is necessary to build a hybrid spam filtering system that makes the use of Naive Bayes and SVM for achieving better accuracy, adaptability and real-time filtering

capability for various types of textual data.

The goal of the study is to create and test such a hybrid model with good benchmark spam data sets for robust and scalable spam detection.

4. Proposed Methodology

The proposed hybrid spam detection framework is based on using both Naive Bayes and Support Vector Machine (SVM) algorithm to achieve better spam detection accuracy and reliability. Training and testing datasets are: SMS Spam Collection and Enron Email Dataset. The pre-processing methods used for the data are stop word removal, stemming, lemmatisation and feature extraction using TF-IDF. Naive Bayes is used to make initial probabilistic classification, and SVM is used to improve predictions when they are not certain by using optimal decision boundaries. The model is tested to ensure efficiency and effectiveness of the model in real time spam detection application based on the metrics of Accuracy, Precision, Recall, F1-Score and Confusion Matrix.

4.1 Proposed Algorithm

The proposed Hybrid classification model is a fusion between Naive Bayes and SVM which overcomes the drawbacks in the conventional spam filtering techniques. Naive Bayes has been used to achieve quick probabilistic classification, while SVM has been used to achieve better classification precision by creating the optimum decision boundaries. This integration not only increases the efficiency of the spam detection process, but also improves the reliability of spam detection.

Methodology Steps:

1. Data Collection and

Preprocessing: Standard datasets such as SMS Spam Collection

Dataset and Enron Email Corpus are collected. The text data is

preprocessed through lowercase conversion, stop-word removal, tokenization, stemming, lemmatization, and dataset splitting.

2. Feature Extraction: TF-IDF vectorization is applied to transform textual messages into numerical feature vectors for machine learning processing.

3. Naive Bayes Classification: A

Multinomial Naive Bayes classifier is trained to perform initial probabilistic spam classification.

4. SVM Classification and Hybrid Fusion: The Support Vector

Machine classifier refines

uncertain predictions, and the outputs of both classifiers are

combined using weighted decision scores for improved accuracy.

5. Performance Evaluation: The hybrid model is evaluated using Accuracy, Precision, Recall, F1-Score, ROC-AUC, and Confusion Matrix.

4.2 Algorithmic Representation

• Hybrid Naive Bayes–SVM Spam Detection:

Input: Text dataset $D =$

$\{X_i, Y_i\}$, where $Y_i \in \{spam, ham\}$

Output:

• Procedure:

1. Perform preprocessing including tokenization, stop-word removal, and stemming.
2. Convert text into TF-IDF feature vectors.
3. Train Naive Bayes classifier to compute spam probabilities.
4. Train SVM classifier to compute decision scores.
5. For each testing message: Compute hybrid score using Naive Bayes probability and SVM score. If score $\geq$ threshold, classify as spam. Otherwise classify as ham
6. Evaluate model using Accuracy, Precision, Recall, and F1-Score.

• Advantages of Proposed Model:

1. Improved classification accuracy
2. Reduction in false positives
3. Better adaptability to evolving spam patterns
4. Computational efficiency
5. Suitable for real-time applications

• **Expected Outcome:** The proposed hybrid model is expected to achieve higher accuracy and robustness compared to standalone classifiers while maintaining efficient real-time spam filtering performance.

4.3 Pseudocode for Hybrid Model

BEGIN

IMPORT required libraries LOAD SMS and Email datasets PREPROCESS text data

- lowercase conversion

- remove stop words

- stemming and lemmatization

- TF-IDF vectorization APPLY Naive Bayes classifier IF confidence is high

ASSIGN class label ELSE

PASS message to SVM classifier APPLY SVM classifier for borderline cases

COMBINE outputs using hybrid decision rule

EVALUATE model using: Accuracy

Precision Recall

F1-Score Confusion Matrix

END

5. Implementation and Results

5.1 System Performance

Several important metrics were used to test the effectiveness of the hybrid spam detection system, such as accuracy, precision, recall, and F1-score. The hybrid model consists of Naive Bayes (NB) and Support vector machine (SVM). The model aims to give more reliable prediction results, particularly for short and long email text.

The platform was tested under concurrent user loads to mimic real-world financial demand, focusing on responsiveness and data integrity.

5.2 Quantitative Performance matrix

For the evaluation of the proposed hybrid spam detection system, the F1 score, accuracy, precision and recall are performed along with other quantitative metrics

commonly used. The metrics can be used to assess the efficacy of the model for the discrimination between spam and non-spam emails in various scenarios. Results of

various models - SVM and Naïve Bayes.

5.3 Result Summary

1. The hybrid system has high accuracy and a balanced performance.
2. NB + SVM is more robust in various spam types.
3. It can be deployed on low resource systems because of fast prediction and lightweight models.
4. The web interface developed using flask is successfully demonstrated with real-time email classification.

5.4 Confusion Matrix

The confusion matrix is used to measure the performance of the spam detection system by showing how accurately emails are classified as spam or non-spam. It includes four main components: True Positives (correctly detected spam), True Negatives (correctly detected genuine emails), False Positives (genuine emails marked as spam), and False Negatives (spam emails marked as genuine).

The hybrid model performed better than individual models because it combines contextual understanding with keyword- based analysis. It achieved high True

Positive and True Negative rates, showing strong accuracy and reliability. At the same time, low False Positive and False Negative values ensured that important emails were not missed and most spam emails were successfully detected. Overall, the confusion matrix shows that the hybrid approach provides a balanced and effective solution for real-world email spam detection.

6. Discussion

The project demonstrated that the hybrid approach using Naive Bayes and SVM is effective for detecting spam or fake emails. Naive Bayes was able to quickly identify obvious spam messages, while SVM performed better in detecting complex patterns and borderline emails. By combining both models, the system achieved a good balance between speed and accuracy.

During testing, it was observed that some spam emails with disguised links or misleading wording could still bypass traditional filters. The use of TF-IDF feature extraction helped highlight important keywords and patterns, improving the model's classification performance. Although the system achieved high accuracy, certain emails were still difficult to classify correctly because context and semantics also play an important role. Overall, the hybrid model provided a reliable and efficient solution for automated spam email detection with minimal human intervention.

7. Challenges Faced

Several challenges were faced during the development of this project. The dataset contained inconsistent formatting, spelling mistakes, special characters, and mixed languages, which made preprocessing difficult. Another challenge was dataset imbalance, as spam emails were fewer compared to normal emails, which could make the model biased toward non-spam predictions.

Training the SVM model also required high computational power because of the large dataset size, leading to longer training time and memory limitations. In addition, combining Naive Bayes and SVM into a hybrid model required careful tuning to ensure that both models contributed fairly to the final prediction.

Despite these challenges, continuous testing and improvements in preprocessing techniques helped enhance the model's performance. These difficulties highlighted the importance of clean data, proper preprocessing, and model tuning in building an effective machine learning system.

8. Future Scope

There is significant scope to further improve the fake email detection system in the future. Advanced machine learning models such as LSTM, BERT, and transformer-based architectures can be used to better understand the context and meaning of emails. Features like URL analysis, attachment scanning, and sender reputation checking can also improve detection accuracy.

The system can be developed into a real-time email scanning tool or browser extension that warns users before opening suspicious emails. Regularly updating the model with new phishing data will help it adapt to evolving cyber threats. In the future, the system can also be deployed on cloud platforms, allowing organizations to integrate it into their cybersecurity infrastructure and making it a more complete security solution.

9. Conclusion

The fake email detection system developed in this project demonstrates how machine learning can effectively help in identifying phishing and fraudulent emails. Through proper preprocessing, feature extraction, and model training, the system achieved reliable performance in detecting malicious emails. The project also showed how automated detection can reduce human effort and improve email security.

At the same time, the study highlights the need for continuous improvement, as cyber attackers constantly develop new techniques to bypass security systems. Although the current model provides a strong foundation, integrating advanced algorithms and adaptive learning methods will be important for long-term effectiveness. Overall, this project offers a practical and scalable approach to improving digital communication security.

10. References

1. Aaron Afan Izang, Oluwabukola F. Ajayi, Omolayo Junaid, Bonaventure Nwigwe, Princewill Onyekachi Onyeka, "Design and Evaluation of Student a Peer-to-Peer Lending Platform to Mitigate Information Asymmetry and Credit Risk", International Information and Engineering Technology Association, Vol. 29, No. 3, June 2024, pp. 885-894
2. Reni Sulastri, Marijn Janssen, "Challenges inclusive in designing Peer-to-peer an (P2P) lending system", Delft University of Technology, July 2023
3. Xinyuan Wei, Jun-ya Gotoh and Stan Uryasev, "Peer-To-Peer Lending: Classification in the Loan Application Process", November 2018 Risks,
4. Kusumawati, Hanafi, "Peer-to-peer lending: An analytical review of research trends prospects", and future Edelweiss Applied Science and Technology, Vol. 8, No. 6, 2024
5. Nisha Arora, Pankaj Deep Kaur, "Case-based reasoning induced optimized investment decision model for retail investors in peer-to peer lending Engineering platforms", Applications of 114376
6. Lyne Imene Souadda, Ahmed Rami Halitim, Billel Benilles, Jose Manuel Oliveira, Patricia Ramos, "Optimizing Credit Risk Prediction for Peer-to-Peer Lending Using Machine Learning",
7. Gurusigaamani Ayyanar Muthulingam, Dinesh Karthik M, Aathi Seshan P, Ganesh Aditya R S, Vijaya Nivas M, "Design and Implementation of a Trust-Based Peer-to-Peer Lending Platform for Hyperlocal Academic Communities", IEEE, September 2025
8. Ashwitha C Thomas, Tanzila Nargis, Abhijith Hegde, Abhishek, Bhumika Shetty, Harsh Bhardwaj, "Peer to Peer Money Lending Using Blockchain", IEEE, February 2025