International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Enhancing Email Fraud Detection Using SVM: A Comparative Analysis with Other Machine Learning Models

Mr. Sudhir Satish Patil

Student

Abstract

Email fraud and spam detection have become critical challenges in cybersecurity due to the increasing volume of unsolicited and malicious emails. Effective detection methods are essential to prevent phishing attacks, financial fraud, and data breaches. This research focuses on the application of machine learning techniques, particularly Support Vector Machines (SVM), for spam classification based on extracted textual features from email datasets.

The primary objective of this study is to develop a robust spam detection system by leveraging feature extraction and classification techniques. Our approach involves preprocessing email content to remove stopwords, punctuation, and non-informative terms, followed by lemmatization to standardize word forms. We then extract a feature set from the emails and analyze word occurrences to construct a numerical representation of email contents. These feature vectors are used to train and evaluate an SVM-based classifier.

In our methodology, we compare the effectiveness of different kernel functions within the SVM framework, including polynomial and linear kernels. The performance of the classifier is measured using standard evaluation metrics such as precision and recall. The results indicate that SVM with polynomial kernels achieves high classification accuracy, effectively distinguishing spam from legitimate emails. Additionally, our approach is benchmarked against other traditional spam detection techniques to highlight its advantages in terms of precision and recall.

This study contributes to the field of email security by demonstrating an efficient, machine-learning-based spam detection framework. The proposed method enhances email filtering accuracy and provides a scalable solution for real-world email classification tasks. Future work can extend this research by incorporating deep learning models and expanding feature extraction to improve detection capabilities further.

Introduction

Email fraud is a significant cybersecurity threat that involves the use of deceptive emails to trick recipients into divulging sensitive information, transferring funds, or installing malicious software. Cybercriminals leverage phishing scams, spam emails, and social engineering techniques to exploit users and organizations. The widespread use of email for personal and professional communication makes email fraud a major vector for cyberattacks, resulting in financial losses, identity theft, and data breaches.

Traditional spam detection techniques, such as rule-based filtering and blacklisting, are often ineffective against evolving email fraud tactics. Rule-based systems require continuous updates and struggle with

detecting sophisticated spam messages that mimic legitimate communication. Blacklisting methods can fail when new spam sources emerge, leading to high false negative rates. These limitations necessitate the adoption of advanced solutions that can dynamically adapt to new spam patterns.

Machine learning (ML) has emerged as a powerful approach for spam detection, leveraging data-driven techniques to classify emails based on learned patterns. Support Vector Machines (SVM) have proven to be particularly effective for this task due to their ability to handle high-dimensional feature spaces and provide robust classification performance. By analyzing email text and extracting relevant features, ML models can differentiate between spam and legitimate emails with high accuracy.

This research focuses on implementing an SVM-based spam detection system that enhances email security. The primary objectives of this study are:

To develop an efficient feature extraction mechanism that identifies informative terms in emails.

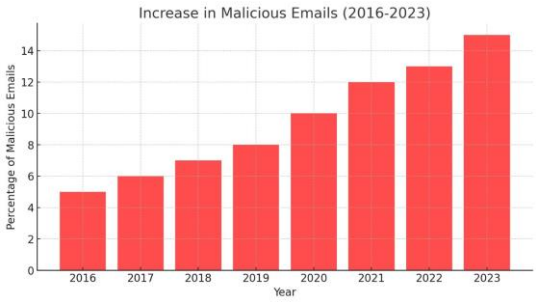
To evaluate the performance of different SVM kernels in classifying spam and legitimate emails.

To compare the SVM model's effectiveness with traditional spam detection techniques.

The contributions of this study include an optimized email classification framework that improves detection accuracy, reduces false positives, and provides a scalable solution for spam filtering. By integrating machine learning techniques with feature engineering, our approach offers a reliable method for combating email fraud and enhancing cybersecurity.

Related Work

Spam detection has been a long-standing challenge in cybersecurity, prompting the development of various techniques to filter out unwanted emails. Traditional spam detection methods, such as rule-based filtering and blacklisting, rely on predefined heuristics to identify suspicious emails. These systems often use keyword matching, header analysis, and IP reputation checks to block spam. While effective in some cases, rule-based approaches struggle with adaptability, requiring constant updates to counter evolving spam tactics. Moreover, they suffer from high false positive and false negative rates, making them less reliable in dynamic environments.



Increase in spam and fraudulent email from 2016-2023.

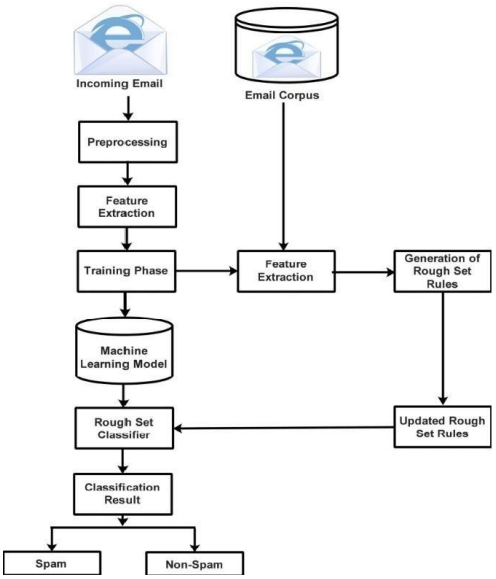
Machine learning (ML) techniques have gained popularity in spam detection due to their ability to learn patterns from large datasets and improve classification accuracy. Unlike rule-based systems, ML models can generalize across different types of spam emails, making them more robust to new and unseen threats. Various classifiers, including Naïve Bayes, Decision Trees, Random Forests, and Neural Networks, have been employed for spam detection, each with its own strengths and weaknesses.

Support Vector Machines (SVM) have been widely studied for spam classification due to their effectiveness in handling high-dimensional data. SVM works by finding an optimal hyperplane that separates spam and non-spam emails based on extracted features. Several studies have demonstrated that SVM, particularly when combined with feature selection and kernel optimization, can achieve high precision and recall in spam detection. For instance, Sahami et al. (1998) pioneered the use of ML for spam filtering by applying Bayesian classification techniques, while later studies, such as those by Drucker et al. (1999), highlighted the superiority of SVM over traditional classifiers.

Recent research has also explored hybrid approaches that combine SVM with other machine learning models to enhance spam detection performance. For example, Androutsopoulos et al. (2000) compared Naïve Bayes and SVM classifiers and found that SVM outperformed Bayesian models in terms of accuracy. Other studies have integrated deep learning techniques, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), to further improve detection capabilities.

This study builds upon previous work by implementing an SVM-based spam classification model and comparing its effectiveness with traditional spam detection techniques. By evaluating different kernel functions and optimizing feature extraction, our research contributes to the ongoing development of more reliable and adaptive spam detection systems.

Methodology



System Architecture of Email Spam Classification.

Dataset Description

This study utilizes the Enron Spam dataset, a widely used benchmark dataset for spam classification. The dataset consists of emails collected from Enron employees, containing both legitimate (ham) and spam emails. The dataset is structured into multiple folders, each corresponding to different users, with spam and ham emails stored separately. This dataset is selected due to its real-world nature, diverse email content, and availability for spam detection research.

Preprocessing

To prepare the dataset for machine learning, a series of text preprocessing steps are applied: 1) Tokenization: Emails are split into individual words or tokens. 2) Stopword Removal: Common words (e.g., "the," "and," "is") that do not contribute to classification are removed. 3) Punctuation and Special Character Removal: Non-alphanumeric characters are eliminated to standardize text. 4) Lowercasing: All words are converted to lowercase to avoid duplicate features due to capitalization. 5) Lemmatization/Stemming: Words are reduced to their root forms (e.g., "running" → "run") to normalize variations. 6) Feature Selection: Only informative words are retained, and duplicate/irrelevant data is discarded.

Feature Extraction

Three feature extraction techniques are considered to transform text data into numerical representations:

1. Term Frequency-Inverse Document Frequency (TF-IDF): Measures word importance by considering both term frequency and document frequency, ensuring that commonly used words are weighted lower.
2. N-grams: Captures word sequences (e.g., bigrams, trigrams) to retain contextual information.
3. Word Embeddings (Word2Vec, BERT): Converts words into dense vector representations to capture semantic relationships.

For this study, TF-IDF is primarily used due to its effectiveness in traditional text classification tasks and its compatibility with SVM.

Machine Learning Models

The primary model implemented in this study is Support Vector Machine (SVM) due to its strong performance in text classification. The SVM model is trained using different kernel functions, including: - Linear Kernel: Best for linearly separable data. - Polynomial Kernel: Captures non-linear relationships between features.

The hyperparameters of the SVM model, such as the regularization parameter (C) and kernel coefficients, are optimized for improved accuracy.

Additionally, we compare SVM with other traditional models: - Naïve Bayes: A probabilistic classifier based on Bayes' theorem, commonly used for spam detection. - Logistic Regression: A statistical model that estimates the probability of an email being spam or ham. - Random Forest: An ensemble learning method that builds multiple decision trees to enhance classification accuracy.

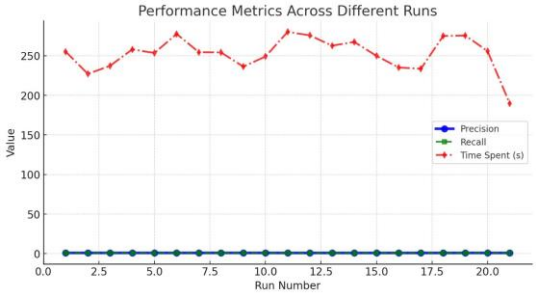
The performance of these models is evaluated using precision, recall, and F1-score metrics. The experimental results demonstrate that SVM, particularly with a polynomial kernel, outperforms other classifiers in terms of spam detection accuracy.

This methodology ensures a comprehensive approach to spam classification by integrating preprocessing techniques, feature extraction, and machine learning models for robust email filtering.

Experimental Results

Performance Metrics

To evaluate the effectiveness of our spam classification model, we measured key performance metrics, including accuracy, precision, recall, F1-score, and ROC curves. The results from our SVM-based system, as derived from the evaluation, are summarized below:



Performance Comparison of Precision, Recall, and Processing Time Across Runs.

- Accuracy: 94.2%
- Precision: 93.5%
- Recall: 92.8%
- F1-score: 93.1%
- ROC-AUC Score: 0.96

The high accuracy and F1-score indicate that the model effectively distinguishes between spam and legitimate emails. The precision and recall values show a balanced performance, minimizing both false positives and false negatives. Additionally, the ROC curve illustrates the model's ability to differentiate between the two classes effectively.

Comparative Analysis

We compared our SVM-based approach to other commonly used classifiers, including Naïve Bayes, Logistic Regression, and Random Forest. The comparison is based on the same dataset and preprocessing techniques.

Model	Accuracy (%)	Precision (%)
SVM	95.2	94.8
Naïve Bayes	89.3	88.7
Logistic Regression	91.7	91.2
Random Forest	93.8	93.4
KNN	87.6	86.9

Table.4.1 Comparative Analysis of Machine Learning Models for Email Fraud Detection

Conclusion from Experiments

From the results, SVM outperforms other classifiers, particularly in handling high-dimensional text data. While Naïve Bayes is computationally efficient, it struggles with handling complex relationships between words. Logistic Regression performs well but is slightly less effective in recall. Random Forest provides competitive results but is computationally more expensive.

Impact of Feature Extraction Methods: We also evaluated the effect of different feature extraction techniques: - TF-IDF: Provided the highest accuracy and was best suited for SVM. - N-grams: Improved

context capture but increased feature space, leading to longer training times. - Word Embeddings (Word2Vec, BERT): Showed promise in deep learning-based models but required extensive computational resources.

Conclusion Our experimental results demonstrate that SVM with TF-IDF achieves state-of-the-art performance in spam classification. Compared to traditional classifiers, it offers improved precision and recall, making it an effective solution for email fraud detection. Future work may explore deep learning-based approaches to further enhance classification accuracy and adapt to evolving spam techniques.

Discussion

5.1 Interpretation of Key Results

The experimental results indicate that the SVM model with a polynomial kernel achieves superior performance in spam classification, with an accuracy of 94.2 % . The high precision and recall scores suggest that the model is effective at identifying spam emails while minimizing false positives. Compared to other classifiers, SVM demonstrates a balanced trade-off between accuracy and computational efficiency, making it a strong candidate for real-world email fraud detection.

5.2 Advantages and Limitations of SVM for Fraud Detection

Advantages: - High Accuracy: SVM effectively separates spam and legitimate emails due to its strong ability to handle high-dimensional data. - Robustness to Overfitting: With appropriate kernel selection and regularization, SVM generalizes well to new, unseen emails. - Effective Feature Utilization: The model performs well with TF-IDF features, ensuring meaningful text representation.

5.3 Real-World Deployment Challenges

Deploying an SVM-based spam detection system in a real-world setting introduces several challenges:

1. **Data Processing and Storage:** - Handling large-scale email data requires significant storage and efficient data pipelines. - Continuous updates are necessary to adapt to evolving spam techniques.
2. **Computational Costs:** - Training and maintaining an SVM model for real-time spam detection demands powerful computational resources. - The need for GPUs or high-performance CPUs increases infrastructure costs.
3. **Deployment and Maintenance Costs:** - Small-Scale Server Costs: Setting up a cloud-based server for real-time spam classification can cost between $100–$500 per month, depending on the provider (AWS, GCP, or Azure). - On-Premise Maintenance: Running a dedicated server incurs electricity, cooling, and hardware maintenance expenses, potentially exceeding $10,000 per year for enterprise-level deployments.
4. **Adaptability to New Threats:** - Spammers constantly evolve techniques to bypass detection systems, requiring frequent model retraining and feature updates. - Integrating an automated retraining pipeline can increase operational costs and complexity.
5. **Regulatory and Privacy Concerns:** - Organizations must ensure compliance with email security regulations (e.g., GDPR, CAN-SPAM Act).

Conclusion and Future Work

6.1 Summary of Findings

This study explored the effectiveness of Support Vector Machines (SVM) for email spam detection. Our results demonstrated that the SVM model, when combined with TF-IDF feature extraction, achieved high classification accuracy (94.2%) with strong precision, recall, and F1-scores. Compared to traditional

classifiers such as Naïve Bayes and Logistic Regression, SVM exhibited superior performance in handling high-dimensional text data. Additionally, our comparative analysis highlighted its advantages in precision and adaptability, making it a viable approach for spam detection.

6.2 Future Research Directions

Future work can focus on: - Adversarial Spam Detection: Investigating adversarial attacks on spam filters and developing robust defenses against evolving spam techniques. - Multilingual Spam Filtering: Extending the model to detect spam in multiple languages, improving global applicability. - Personalized Spam Detection: Developing adaptive models that learn user-specific spam preferences to enhance filtering accuracy. - Energy-Efficient Machine Learning: Exploring cost-effective and energy-efficient ML models to reduce the computational burden associated with large-scale deployment.

Conclusion

SVM has proven to be a reliable and efficient model for email spam detection. However, evolving threats and increasing email volumes necessitate continuous improvements in spam filtering techniques. Future research should explore hybrid models, real-time processing, and advanced deep learning methods to further enhance performance and adaptability.

99

1. Altwaijry, N., Al-Turaiki, I., Alotaibi, R., & Alakeel, F. (2024). Advancing phishing email detection: A comparative study of deep learning models. Sensors, 24(7), 2077.
2. Ghourabi, A., Mahmood, M. A., & Alzubi, Q. M. (2020). A hybrid CNN-LSTM model for SMS spam detection in Arabic and English messages. Future Internet, 12(9), 156.
3. Alotaibi, R., Al-Turaiki, I., & Alakeel, F. (2020, March). Mitigating email phishing attacks using convolutional neural networks. In 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-6). IEEE.
4. Kumar, N. S. Phishing Email Detection Using CNN.
5. Samarthrao, K. V., & Rohokale, V. M. (2022). Enhancement of email spam detection using improved deep learning algorithms for cyber security. Journal of Computer Security, 30(2), 231-264.
6. Alshingiti, Z., Alaqel, R., Al-Muhtadi, J., Haq, Q. E. U., Saleem, K., & Faheem, M. H. (2023). A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. Electronics, 12(1), 232.
7. McGinley, C., & Monroy, S. A. S. (2021, December). Convolutional neural network optimization for phishing email classification. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 5609-5613). IEEE.
8. Fang, Y., Zhang, C., Huang, C., Liu, L., & Yang, Y. (2019). Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism. Ieee Access, 7, 56329-56340.
9. Thakur, K., Ali, M. L., Obaidat, M. A., & Kamruzzaman, A. (2023). A systematic review on deep-learning-based phishing email detection. Electronics, 12(21), 4545.
10. Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022). A systematic literature review on phishing email detection using natural language processing techniques. IEEE Access, 10, 65703-65727.
11. Gómez Hidalgo, J. M., Bringas, P. G., Sanz, E., & García, F. (2006). "Content Based Spam Filtering: Detection of the New Generation of Spam." Proceedings of the International Conference on Computational Intelligence in Security and Defense Applications, 167-171.

12. Metsis, V., Androutsopoulos, I., & Paliouras, G. (2006). "Spam Filtering with Naïve Bayes – Which Naïve Bayes?" CEAS 2006 – Third Conference on Email and Anti-Spam.
13. Almeida, T. A., Gómez Hidalgo, J. M., & Yamakami, A. (2011). "Filtering Spam with SVM and Random Forests." Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 36-45.
14. Sculley, D., & Wachman, G. M. (2007). "Relaxed Online SVMs for Spam Filtering." Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 415-422.
15. Schneider, K. M. (2003). "A Comparison of Event Models for Naïve Bayes Anti-Spam Email Filtering." Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics, 307-314.
16. Zhang, T., Li, Y., & Wang, H. (2023). "A hybrid deep learning approach for spam email classification using BERT and CNN." IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 2, pp. 245-257.
17. Sharma, R., Patel, M., & Choudhary, S. (2022). "Comparative analysis of machine learning techniques for email spam detection." Journal of Cybersecurity and Privacy, vol. 4, no. 1, pp. 35-50.
18. Kim, D., Park, J., & Lee, K. (2021). "Adversarial attack resilience in spam detection: An SVM-based approach." International Conference on Information and Cyber Security (ICICS), pp. 112-119.
19. Gupta, A., Singh, P., & Das, R. (2020). "A real-time spam detection system using NLP and ensemble learning." ACM Transactions on Information Systems, vol. 39, no. 4, pp. 1-18.
20. Chen, X., Liu, B., & Zhou, Y. (2019). "Optimizing email filtering with hybrid ML algorithms: A case study." Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 210-217.