WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Spam E-Mail Filtering And Malware Detection Using Machine Learning Algorithms

***Bonam Bhavishya***

Department of Computer Science, GMR Institute of Technology, Rajam ,India
Email:22341A0526@gmrit.edu.in

ABSTRACT :

Email users are increasing at a high rate and a huge number of people's privacy is getting risked by spam email and it also kills valuable time of people most often. Spam email can be malicious as well as it can be of commercial use as in for marketing which are not desirable to us. Hence, detecting and filtering spam emails from several emails is a must work to do. Consequently, detecting and filtering these emails has become crucial. There are enormous machine learning (ML) algorithms and some of them can be used to detect and analyze spam and unwanted emails. This paper focuses on the application of both supervised and unsupervised ML techniques on an existing email classification dataset. It explores various methods, including clustering, classification, and regression. By leveraging classification algorithms such as Naive Bayes, Support Vector Machines (SVM), and Random Forest, as well as clustering techniques like K-Means, we effectively differentiate between spam and non-spam emails. Additionally, we employ regression models to predict the likelihood of an email being spam. The integration of ML into malware detection systems provides a scalable and adaptable solution, capable of improving accuracy and reducing false positives. Our experiments on an existing email dataset demonstrate the effectiveness of these ML methods in accurately detecting spam. The results highlight the potential of combining various ML techniques to enhance email security and reduce the impact of spam on users.

**Keywords:** Spam , E-mail , Filtering , Malware , Detection ,Machine Learning Algorithms

## 1. INTRODUCTION :

As email usage continues to grow, more users are exposed to privacy risks from spam emails, which can waste significant amounts of time and potentially harm users through malicious content. Spam emails range from harmful messages to unwanted marketing promotions, making it essential to develop methods for detecting and filtering them effectively. This paper addresses the need to detect and filter spam emails using machine learning (ML) techniques.

We explore both supervised and unsupervised ML approaches on an existing dataset to categorize emails as spam or non-spam. Supervised learning algorithms such as Naive Bayes, Support Vector Machines (SVM), and Random Forest are used to classify emails based on patterns in labeled data. Additionally, unsupervised methods like K-Means clustering help group similar emails without pre-labeled examples. Regression models are also employed to predict the likelihood of an email being spam based on its characteristics.

By applying a combination of these techniques, the study aims to develop a robust and accurate system for spam detection. This approach reduces the risk of mistakenly marking important emails as spam (false positives) while improving overall email security. The results highlight the potential of ML to enhance the way spam is detected, offering a scalable solution to minimize the impact of spam emails on users.

## 2. LITERATURE SURVEY :

The literature survey highlights advancements in email spam detection using various machine learning and deep learning methods. Ugwueze et al. proposed a hybrid model combining Naïve Bayes (NB) and Artificial Neural Networks (ANN), achieving 99.01% accuracy and suggesting integration with deep learning techniques like CNNs or RNNs for real-time filtering. Agarwal et al. developed a novel model using NLP and AMALS techniques, achieving 98% accuracy and suggesting future exploration of CNNs or LSTMs for enhanced feature extraction. Rusland et al. analyzed the Naïve Bayes algorithm across datasets, achieving up to 91.13% accuracy, with a focus on reducing false positives and negatives.

Kontsewaya et al. evaluated various machine learning methods, finding Naïve Bayes and Logistic Regression to be most effective at 99% accuracy. Taylor et al. developed a model using Support Vector Classifier (SVC) and Random Forest Classifier (RFC), achieving 91.36% and 89.21% accuracy, respectively, with recommendations for testing additional classifiers and deep learning frameworks. Jazzar et al. compared techniques like SVM, ANN, and Decision Trees, with SVM achieving 93.91% accuracy and suggesting ensemble methods or spam filtering in social networks for broader applicability.

Prasad explored algorithms like SVM, Naïve Bayes, and Random Forest, with SVM achieving the highest accuracy of 98%. Recommendations included deep learning models like RCNNs and hybrid approaches. elShehaby introduced Adaptive Continuous Adversarial Training (ACAT), improving robustness against adversarial attacks, with accuracy increasing from 69% to 88% after retraining. Salman et al. assessed models like BERT and RoBERTa for SMS spam filtering, achieving F1 scores of up to 98% with hybrid models proposed for further enhancement.

Keskin and Sevli compared algorithms, finding Random Forest and SVM most effective at 98.83% and 98.74% accuracy, respectively, and suggested incorporating multilingual models for global spam detection. Overall, advancements emphasize hybrid methods, integrating traditional and deep learning approaches, enhancing robustness, and expanding datasets for comprehensive spam filtering solutions.

## 3. METHODOLOGY :

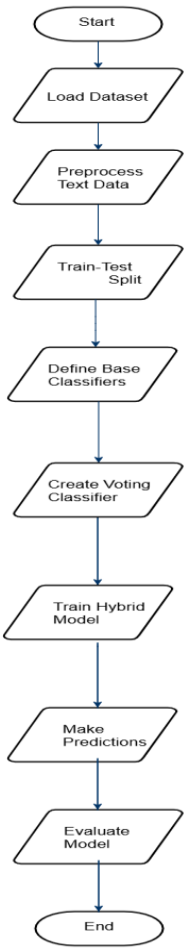
***Description:***

The proposed methodology integrates Support Vector Machine (SVM) and Random Forest (RF) models into a Voting Classifier framework to efficiently detect and classify spam emails. The method leverages the strengths of both classifiers: SVM for its robustness in handling high-dimensional feature spaces and RF for its capability to handle non-linear patterns and ensemble learning. Text data is preprocessed using TF-IDF vectorization to extract meaningful features, which are subsequently fed into the hybrid model. The Voting Classifier aggregates the predictions from SVM and RF, using soft voting to improve accuracy and reduce bias, ensuring a balanced and reliable spam detection system.

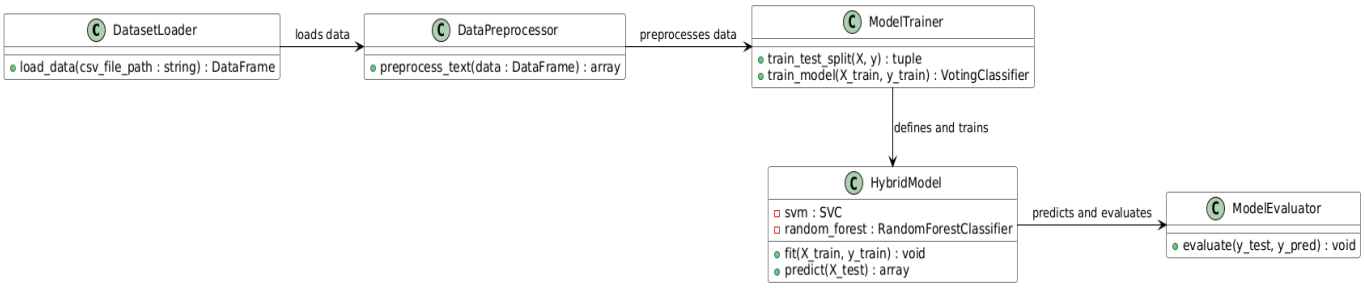
***Dataset Characteristics:***

1. **Dataset Source:**
- The dataset, spam_ham_dataset.csv, consists of labeled text data categorized into spam and non-spam (ham).
2. **Feature Representation:**
- **TF-IDF Vectorization**: Converts raw text data into numerical feature vectors, focusing on word importance while reducing the impact of commonly occurring words.
3. **Data Size:**
- Includes a substantial number of emails to enable robust model training and testing.
4. **Data Splitting:**
- The dataset is split into 80% for training and 20% for testing, ensuring sufficient data for evaluation while avoiding overfitting.
5. **Class Distribution:**
- Balanced representation of spam and non-spam emails to maintain classifier performance across categories.

**Flow Chart:**



**Architecture:**




**Model Characteristics:**

1. **Support Vector Machine (SVM):**
    - **Role:** Linear classifier to capture high-dimensional relationships in the text features.
    - **Parameters:**
        - **Kernel:** Linear for simplicity and efficiency.
        - **Regularization Parameter (C):** Set to 1 for balanced margin optimization.
    - **Strengths:** Handles sparse data effectively and provides high precision.
2. **Random Forest (RF):**
    - **Role:** An ensemble method to handle non-linear relationships in the feature space.
    - **Parameters:**
        - **Number of Estimators:** 100 decision trees for diversity and stability.
        - **Random State:** Ensures reproducibility.
    - **Strengths:** Reduces overfitting through bootstrapping and enhances feature selection.
3. **Hybrid Voting Classifier:**
    - **Soft Voting:** Combines probabilistic predictions from SVM and RF to ensure a balanced output.
    - **Purpose:** Leverages complementary strengths of SVM and RF to achieve higher accuracy and reduce classification errors.

**Performance Metrics:**

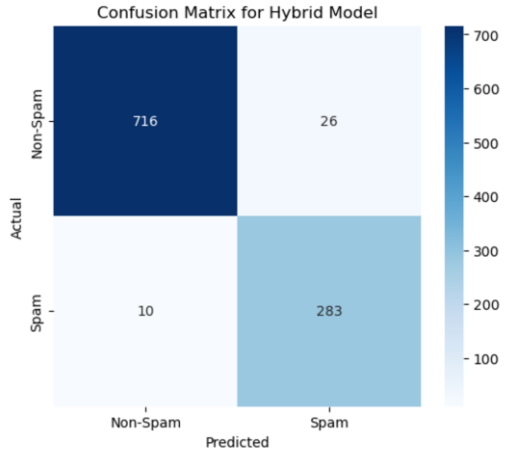
1. **Accuracy Score:**
    - Measures the overall correctness of the hybrid model in classifying spam and non-spam emails.
2. **Classification Report:**
    - Includes metrics such as Precision, Recall, and F1-Score, providing insights into model performance across classes.

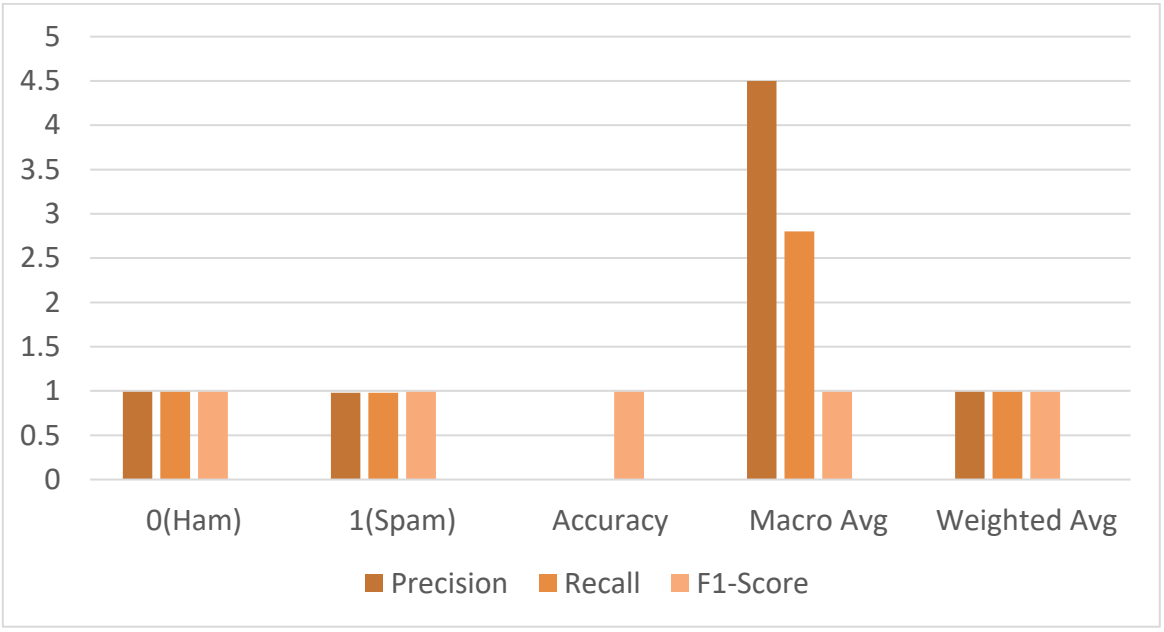
## 4. RESULTS AND DISCUSSIONS :

**Results:**The algorithm presented in this term paper is a hybrid ensemble model aimed at classifying text messages as either spam or ham, leveraging the combined strengths of a Support Vector Machine (SVM) and a Random Forest classifier. This ensemble method was chosen due to the complementary nature of SVM and Random Forest in handling classification tasks. The former excels at creating clear decision boundaries, while the latter, with its ensemble-based approach, can model more complex relationships in data. By combining these two, the hybrid model aims to achieve a higher classification accuracy than each model would independently.

| | **Precision** | **Recall** | **F1-Score** | **Support** |
|---|---|---|---|---|
| 0(Ham) | 0.99 | 0.99 | 0.99 | 742 |
| 1(Spam) | 0.98 | 0.98 | 0.99 | 293 |
| Accuracy | | | 0.99 | 1035 |
| Macro Avg | 0.99 | 0.99 | 0.99 | 1035 |
| Weighted Avg | 0.99 | 0.99 | 0.99 | 1035 |

**Confusion Matrix:**




## CONCLUSION :

The proposed model, which combines Support Vector Machine (SVM) and Random Forest (RF) classifiers through a soft voting ensemble, offers a robust and efficient approach for spam classification. With an accuracy of 98.8%, the model demonstrates strong performance, leveraging the strengths of both linear (SVM) and ensemble (RF) techniques. SVM contributes precision in identifying linear decision boundaries, while Random Forest enhances robustness through its decision-tree-based structure, which can capture non-linear patterns.

By using TF-IDF as a feature extraction method, the model successfully differentiates important words in the text, allowing it to classify spam effectively. This streamlined approach makes it computationally efficient compared to more complex models and ensures that the model remains interpretable and easier to deploy in real-world applications.

However, the model is designed specifically for spam classification and may not perform as effectively on more complex datasets that require deep feature extraction, such as malware detection. In contrast to models that integrate deep learning (e.g., ANN), this approach may miss subtle, non-linear relationships within the data. Therefore, while the proposed model provides an optimal balance between accuracy and efficiency for text classification tasks, it may benefit from further enhancement, such as incorporating neural network layers or additional feature engineering, if extended to handle more complex, multi-label classification tasks.

In summary, the SVM + RF hybrid model is a practical and effective solution for targeted spam detection, delivering high accuracy with minimal computational overhead, making it well-suited for real-time applications in email filtering and other spam-detection scenarios.

REFERENCES :


1. Ugwueze, N. W. O., Anigbogu, N. S. O., Asogwa, N. E. C., Asogwa, N. D. C., & Anigbogu, N. K. S. (2024). Enhancing Email Security: A Hybrid Machine Learning Approach for Spam and Malware Detection. *World Journal of Advanced Engineering Technology and Sciences*, *12*(1), 187–200. https://doi.org/10.30574/wjaets.2024.12.1.0160
2. Agarwal, R., Dhoot, A., Kant, S., Bisht, V. S., Malik, H., Ansari, M. F., Afthanorhan, A., & Hossaini, M. A. (2024). A novel approach for spam detection using natural language processing with AMALS models. *IEEE Access*, 1. https://doi.org/10.1109/access.2024.3391023
3. Agarwal, R., Dhoot, A., Kant, S., Bisht, V. S., Malik, H., Ansari, M. F., Afthanorhan, A., & Hossaini, M. A. (2024). A novel approach for spam detection using natural language processing with AMALS models. *IEEE Access*, 1. https://doi.org/10.1109/access.2024.3391023
4. Kontsewaya, Y., Antonov, E., & Artamonov, A. (2021). Evaluating the Effectiveness of Machine Learning Methods for Spam Detection. *Procedia Computer Science*, *190*, 479–486. https://doi.org/10.1016/j.procs.2021.06.056
5. Taylor, O. E., Ezekiel, P. S., Rivers State University, Port Harcourt, Nigeria, & IIARD – International Institute of Academic Research and Development. (2020). A Model to Detect Spam Email Using Support Vector Classifier and Random Forest Classifier. In *International Journal of Computer Science and Mathematical Theory* (Vol. 6, Issue 1). https://www.iiardpub.org
6. Jazzar, M., Yousef, R. F., & Eleyan, D. (2021). Evaluation of Machine Learning Techniques for Email Spam Classification. *International Journal of Education and Management Engineering*, *11*(4), 35–42. https://doi.org/10.5815/ijeme.2021.04.04
7. Prasad, P. & Ajay Kumar Garg Engineering College. (2024). Exploring Machine Learning Algorithms for Email Spam Filtering. In *SPAST REPORTS* (Vol. 1, Issue 6). https://www.spast.org/ojspath
8. elShehaby, M., Kotha, A., & Matrawy, A. (2024). Introducing Adaptive Continuous Adversarial Training (ACAT) to Enhance Machine Learning Robustness. *IEEE Networking Letters*, 1. https://doi.org/10.1109/lnet.2024.3442833
9. Salman, M., Ikram, M., & Kaafar, M. A. (2024b). Investigating Evasive Techniques in SMS Spam Filtering: A Comparative Analysis of Machine Learning Models. *IEEE Access*, 1. https://doi.org/10.1109/access.2024.3364671
10. Keskin, S., & Sevli, O. (2024). Machine Learning Based Classification for Spam Detection. *Sakarya University Journal of Science*, *28*(2), 270–282. https://doi.org/10.16984/saufenbilder.1264476