Enhancing Email Security: A Machine Learning Approach for Phishing Detection

Department of Computer Science, Stephen F. Austin State University

[1]Adwaith Atholi Thiruvoth, Faculty Mentor: [2*]Dr. Pushkar Ogale

[1]Email: atholitha@jacks.sfasu.edu, [2*]Email: ogalep@sfasu.edu

ABSTRACT

Email phishing is a pressing cybersecurity challenge that requires efficient detection methods. Emails that look legitimate lead users to malicious sites. This project aims to develop a machine learning-driven email classification system. A comparative study of classical machine learning techniques like Random Forest, Naive Bayes, Decision Tree, SVM, and gradient boosting regression trees was conducted, and it was found to be very successful in achieving This study makes use of various statistical methods, classification algorithms to develop a user-friendly graphical interface (GUI) for seamless email classification. This system automatically fetches the mailbox file associated with the current user in a Linux environment and the GUI enables users to predict the labels of their emails and view the results interactively. This project helps to strengthen email security by providing a convenient tool for phishing email identification, thereby enhancing defense against cyber threats.

INTRODUCTION

In an era dominated by digital communication, email is still a widely used tool for personal and professional communication. However, it also serves as a medium for cyber attacks, especially phishing attacks. Phishing emails trick individuals into revealing sensitive information, such as passwords and personally identifiable information, posing risk to individuals and organizations. Traditional email filtering methods struggle to keep pace with the evolving attack tactics. Hence, there is a need to use machine learning to improve email classification and phishing detection. This project addresses this challenge by developing an email classification system with the help of machine learning. Leveraging labeled datasets, various algorithms—Naive Bayes, Decision Tree, Random Forest, Gradient Boosting Regression Trees, and Support Vector Machine—are trained and evaluated for their effectiveness. Feature extraction methods like Term Frequency-Inverse Document Frequency (TF-IDF) are used to transform raw email text to a suitable numerical format. The project culminated in a user-friendly GUI, which enables seamless interaction with the email classification system.

Figure 1. Word Cloud of Phishing Emails

Fig 1 is a word cloud visualization that represents a graphical summary of the most frequently occurring words in phishing emails. Each word's size corresponds to its frequency in the text, with larger words indicating higher occurrence. This visualization provides insights into the common themes and language patterns found in phishing emails, aiding in understanding the characteristics of such malicious content.

RESEARCH OBJECTIVE

The objective of the paper is to develop a machine learning-based system to classify phishing and legitimate emails. The research aims to compare supervised learning algorithms, including Naive Bayes, Decision Tree, Random Forest, Gradient Boosting Regression Tree, and SVM in order to identify the most effective approach for email classification based on the requirements and performance metrics such as accuracy, precision, recall, and F1 score.

METHODOLOGY

The methodology employed in this research involves the following steps:

1. Dataset Acquisition: Collect a diverse dataset of labeled email samples consisting of both legitimate and phishing emails.
2. Data Preprocessing: Cleanse the data by removing special characters, standardizing text formats, and handling missing values.
3. Feature Extraction: Utilize the TF-IDF (Term Frequency-Inverse Document Frequency) technique to extract features from the email text data. This process involves tokenizing the text, computing the term frequency of each token within each email, and weighing the importance of each token based on its frequency across all emails.
4. Model Training: Train multiple machine learning models, using the preprocessed dataset.
5. Model Evaluation: Evaluate the performance of each trained model using metrics such as accuracy, precision, recall, and F1 score. Assess the models' ability to accurately classify emails and detect phishing attempts.
6. Integration and Deployment: Integrate the best-performing model into a unified email classification system. Develop a user-friendly GUI to enable users to interact with the system easily.
7. Testing and Validation: Conduct extensive testing and validation of the email classification system to ensure its effectiveness and reliability. Validate the system's performance against known phishing email samples and real-world scenarios.
8. Performance Analysis: Analyze the performance of different machine learning algorithms and feature extraction techniques. Compare the strengths and weaknesses of each model to identify opportunities for improvement and optimization.

RESULTS

The comparison of the algorithms used in this project was based on their performance in accurately classifying emails as either legitimate or phishing. During the training process, we split the dataset such that 80% of it is used for training the remaining is used to test the algorithm.

By evaluating metrics such as accuracy, precision, recall, and F1 score, we assess the effectiveness of each algorithm in distinguishing between the two classes of emails. This comparative analysis aims to identify the most suitable algorithm for developing a robust email classification system capable of mitigating the risks posed by phishing attacks and enhancing overall cybersecurity.

Results indicated that each algorithm achieved high performance in classifying emails as phishing and legitimate. SVM and Random Forest produced the best results with an F1 score of 97.75% and 96.73%.

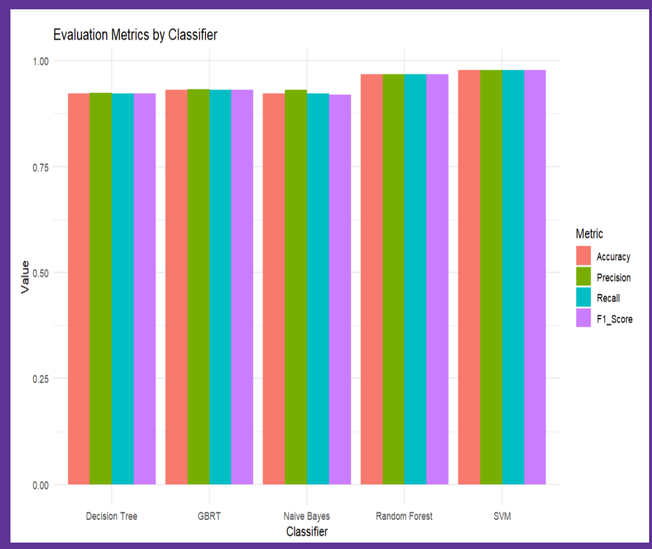

Figure 2. Comparison of Evaluation Metrics by Classifier

CONCLUSION

The findings of this study highlight the effectiveness of different supervised machine-learning algorithms. SVM exhibited the highest overall efficiency followed by the Random Forest model. Even though both these models are suitable candidates to build an email classification system, considering factors such as computational efficiency and scalability, it was determined that Random Forest was the optimal choice for the task, exhibiting high levels of accuracy, precision, recall, and F1 score.

The selection of the algorithm was followed by developing an email classifier system integrated with Graphical User Interface, enabling seamless input of email text for classification and presenting real-time classification results.

FUTURE DIRECTIONS

While this study provided insights into various supervised learning algorithms for email classification, there is room for future research and development.

1. Enhanced Feature Engineering: Explore the advanced feature engineering techniques to extract more discriminative features from email text.
2. Real-time Detection and Response: Improving security by developing a real-time monitoring tool that continuously scans incoming emails and notifies users upon detecting a phishing email.
3. Platform Compatibility: Adapting this tool to work within a Windows environment, providing broader accessibility and integration with common email clients and server configuration.

REFERENCES

1. Harikrishnan, N. B., Vinayakumar, R., & Soman, K. P. (2018, March). A machine learning approach towards phishing email detection. In Proceedings of the anti-phishing pilot at ACM international workshop on security and privacy analytics (IWSPA AP) (Vol. 2013, pp. 455-468).
2. Zuhair, H., Selamat, A., & Salleh, M. (2016). Feature selection for phishing detection: a review of research. International Journal of Intelligent Systems Technologies and Applications, 15(2), 147-162.

ACKNOWLEDGEMENTS

I extend my heartfelt appreciation to the organizers of the ACM International Workshop on Security and Privacy Analytics for granting me the valuable opportunity to share the outcomes of this research.

I would like to express my gratitude to my supervisor, Dr. Pushkar Ogale for his unwavering guidance, expertise, and encouragement throughout the duration of this project.