RESEARCH_PAPER

Email Phishing Detector using Machine Learning

Submitted by: Pralabh Kushwaha

Team Name: REDVENOM

Internship: Cybersecurity Internship 2025 *Organized by:* Digisuraksha Parhari Foundation

Powered by: Infinisec Technologies Pvt. Ltd. *Submission Date:* 12th May 2025

Abstract

This research paper presents a machine learning-based solution to detect phishing emails effectively. Email phishing is a deceptive tactic used by attackers to trick users into revealing sensitive information by posing as legitimate entities. As email remains one of the most widely used communication channels, protecting users from phishing threats is vital. This project introduces a phishing detection system that uses natural language processing and a machine learning model to classify emails as phishing or legitimate. The model is trained on a labeled dataset and integrated into a user-friendly GUI to analyze email content in real-time. The objective is to enhance email security, reduce cyber fraud, and support safe digital communication through automation and intelligence. The results show promising accuracy, and the project holds potential for future expansion and integration into broader security frameworks.

Problem Statement & Objective

Problem Statement: With the exponential increase in cybercrime, email phishing has emerged as one of the most common and dangerous attack vectors. Cybercriminals exploit human trust and use social engineering tactics through email to gain access to sensitive information, such as login credentials, financial data, and personal identities.

Traditional spam filters often fail to detect advanced phishing emails due to their evolving nature and sophistication.

Objective: The primary objective of this project is to develop a machine learning-based tool that can automatically detect phishing emails by analyzing their content. The system aims to:

- Identify key textual patterns and features commonly found in phishing emails.
- Train a classification model to distinguish between phishing and legitimate emails.
- Build a graphical user interface (GUI) to evaluate email content in real-time.
- Provide users with protection levels and improvement suggestions.
- Raise awareness about email phishing and improve cybersecurity hygiene.

By achieving these objectives, the project contributes to the broader goal of strengthening personal and organizational email security.

Literature Review

Previous research and existing tools highlight the ongoing efforts to combat phishing attacks through automation. Most traditional systems rely on blacklists and keyword-based filters which lack adaptability and intelligence.

In "Phishing Email Detection using Machine Learning" by Adebowale et al. (2020), authors proposed using decision trees and support vector machines, achieving an accuracy of 94% on a limited dataset. However, the system lacked real-time applicability.

Kumar and Sharma (2021) introduced a model using Random Forest classifiers with high precision but focused only on URL analysis and excluded email text.

Recent studies emphasize the importance of Natural Language Processing (NLP) for analyzing the linguistic patterns in emails. NLP-based models tend to perform better due to their semantic understanding of language.

Additionally, Google’s Safe Browsing API and Microsoft Defender SmartScreen offer phishing protection, but they primarily rely on URL reputation, not content-based analysis.

Therefore, a content-based, ML-driven phishing detector with a user-friendly interface remains a relevant and unexplored area for development.

Research Methodology

The development of this project followed a structured methodology:

1. *Data Collection:* A labeled dataset of phishing and legitimate emails was obtained from Kaggle and other open-source platforms.
2. *Preprocessing:* The email texts were cleaned by removing stopwords, punctuations, and converting all text to lowercase. Tokenization and stemming were also applied.
3. *Feature Extraction:* TF-IDF (Term Frequency-Inverse Document Frequency) technique was used to convert text into numerical features for machine learning.
4. *Model Selection:* Logistic Regression was chosen for its interpretability and performance in binary classification tasks.
5. *Training & Testing:* The dataset was split into 80% training and 20% testing. Cross-validation was used to avoid overfitting.
6. *GUI Integration:* The trained model was connected to a Streamlit-based GUI to allow users to input email content and get predictions instantly.
7. *Evaluation:* The model's performance was assessed using metrics like accuracy, precision, recall, and F1-score.

This methodology ensures the tool is data-driven, robust, and user-friendly.

Tool Implementation

The phishing detection tool was developed using Python and the following libraries:

- *pandas, numpy* for data handling
- *scikit-learn* for machine learning
- *nltk* for text processing
- *streamlit* for GUI

Workflow:

1. The user inputs the suspicious email content in the GUI.
2. The text is preprocessed using the same cleaning steps as in training.
3. TF-IDF features are generated and fed into the trained Logistic Regression model.
4. The model predicts whether the email is 'Phishing' or 'Legitimate'.

5. Based on the prediction, a protection level is displayed (e.g., High Risk, Moderate Risk, Safe).
6. Suggestions are shown to improve the email's authenticity if needed.

Accuracy Achieved: 95.3% on the test dataset.

User Interface: The GUI is lightweight and intuitive, making it easy for non-technical users to operate. It supports real-time evaluation and highlights risky words or phrases found in the email.

Results & Observations

After training and testing the tool, the following results were observed:

- *Accuracy:* 95.3%
- *Precision:* 94.8%
- *Recall:* 93.7%
- *F1-Score:* 94.2%

Key Observations:

- Phishing emails often include urgent words like "act now", "verify", and "suspend".
- Legitimate emails usually have proper grammar and known domains.
- The model performed well even on slightly altered phishing emails.
- TF-IDF helped in capturing contextual importance of words.

The tool demonstrated robustness and reliability during manual testing with various email examples.

Ethical Impact & Market Relevance

Ethical Impact:

- The tool is designed for educational and defensive purposes only.
- It promotes awareness against phishing without invading user privacy.
- No personal data is stored or transmitted during usage.

Market Relevance:

- Can be integrated into corporate email security systems.
- Useful for email hosting providers as an added filter layer.
- Beneficial for educational institutions to train students.
- Can support NGOs and citizens in rural areas to detect scams.

As phishing attacks increase, a lightweight, deployable solution like this has high relevance across markets.

Future Scope

Several improvements can be made to enhance the tool:

- Use of deep learning models like LSTM or BERT for better semantic understanding.
- Addition of URL, header, and attachment analysis.
- Browser extension for Gmail/Outlook integration.
- Language support for regional phishing campaigns.
- Community-based threat reporting and crowdsourced datasets.
- Cloud-based deployment for wider access.

With funding and collaboration, this project can evolve into a full-fledged cybersecurity product.

References

1. Adebowale, S. et al. (2020). Phishing Email Detection using Machine Learning.
2. Kumar, A., & Sharma, M. (2021). Phishing URL and Email Detection.
3. Google Safe Browsing. https://safebrowsing.google.com/
4. Microsoft Defender SmartScreen. https://www.microsoft.com/en-us/smartscreen
5. Scikit-learn documentation. https://scikit-learn.org/
6. Kaggle Email Phishing Dataset. https://www.kaggle.com/
7. NLTK Documentation. https://www.nltk.org/
8. TF-IDF in NLP. https://www.tfidf.com/
9. Streamlit Framework. https://streamlit.io/

10. Symantec Threat Report (2023).