



RESEARCH PAPER – AI-Powered Phishing Email Detector

1. Title



AI-Powered Phishing Email Detector



Internship Program: [Digi Suraksha Cyber Security Internship 2025](#)



Submitted By: Bhavesh Kerkar & Devraj Pujari



Submission Date: 12th May 2025

2. Abstract

Phishing attacks are one of the most common cybersecurity threats faced by individuals and organizations. These attacks aim to deceive users into revealing sensitive information such as passwords, bank details, and personal data. This project presents an AI-powered phishing email detection tool that uses machine learning techniques to classify emails as either phishing or legitimate based on their textual content. The tool uses a supervised learning approach with text preprocessing, TF-IDF vectorization, and a Naive Bayes classifier. The system demonstrates high accuracy on a public dataset and serves as a lightweight solution for phishing detection. This project aims to support email users by increasing awareness and offering a basic level of security through automation.

3. Problem Statement & Objective

The goal of this project is to reduce the number of successful phishing attacks by using an automated machine learning model that detects phishing emails. Traditional spam filters often fail to identify sophisticated phishing attempts. This project seeks to create a lightweight, intelligent system that improves email security through natural language processing and classification techniques.

4. Literature Review

Several research studies have proposed machine learning approaches to detect phishing. Most solutions rely on lexical features, URL analysis, and metadata. Some use natural language processing (NLP) to analyze the email body, subject lines, and sender details. Naive Bayes, Logistic Regression, Random Forests, and Deep Learning models have all shown promise. However, most require complex datasets or high computational resources.

Our model focuses on content analysis using TF-IDF and a Naive Bayes classifier, balancing accuracy with simplicity and speed.

5. Research Methodology

1. **Data Collection:** A labeled dataset containing phishing and legitimate email texts was downloaded from Kaggle.
 2. **Data Preprocessing:** Converted all text to lowercase, removed punctuation, and tokenized.
 3. **Vectorization:** Used TF-IDF to convert text into numerical vectors.
 4. **Model Training:** Trained a Naive Bayes classifier.
 5. **Testing:** Used an 80-20 train-test split for validation.
-

6. Tool Implementation

The tool is implemented in Python. The user provides email content via a command-line interface, and the model predicts whether it's phishing or legitimate. Libraries used include `pandas`, `scikit-learn`, and `nltk`. The GitHub repository contains the full source code, `requirements.txt`, and setup instructions. The final trained model offers around 95% accuracy on the test data.

7. Results & Observations

- **Model Used:** Naive Bayes
 - **Accuracy Achieved:** ~95%
 - **Precision & Recall:** Balanced performance on both phishing and legitimate classes.
 - **Observation:** Model performs well on known patterns but may need retraining with evolving phishing techniques.
-

8. Ethical Impact & Market Relevance

The tool is ethical and open-source. It is intended for educational and personal use. It does not harvest or send data. This kind of tool has practical relevance for individuals, small businesses, and email client developers who want to integrate ML-based spam detection.

9. Future Scope

- Add browser-based interface (Streamlit)
 - Integrate real-time email fetching (IMAP)
 - Train on larger, real-world datasets
 - Use deep learning (LSTM or BERT) for improved detection
-

10. Screenshots

1. Dataset Preview

1	EmailText	Label
2	Urgent! Your account has been suspended.	1
3	Meeting is scheduled at 4 PM today.	0
4	You won a lottery! Click the link to claim.	1
5	Here is the monthly report you requested.	0
6	Please verify your account information.	1
7	Reminder: Your subscription is about to expire.	0
8	Get your free gift card now! Click here.	1
9	Your package has been shipped.	0
10	Important: Your payment details need updating.	1
11	Company holiday party details inside.	0
12	Claim your prize now before it's too late.	1
13	Join our webinar on digital marketing tomorrow.	0
14	You've received a new message from HR.	0
15	Your credit card is about to expire. Update now.	1
16	Special offer just for you! Get 50% off.	1
17	New job opportunity at XYZ company.	0
18	You've been selected for a special promotion!	1

2. TF-IDF Vector Output /Preprocessing Output:

1	EmailText	Label
2	Urgent! Your account has been suspended.	1
3	Meeting is scheduled at 4 PM today.	0
4	You won a lottery! Click the link to claim.	1
5	Here is the monthly report you requested.	0
6	Please verify your account information.	1
7	Reminder: Your subscription is about to expire.	0

3. Model Training Console Output

```
Accuracy: 0.3333333333333333

Report:
              precision    recall  f1-score   support

     0       0.20      1.00      0.33         1
     1       1.00      0.20      0.33         5

 accuracy          0.33         6
 macro avg          0.60      0.60      0.33         6
weighted avg          0.87      0.33      0.33         6
```

4. GUI Interface (Home Screen)

Phishing Email Detection

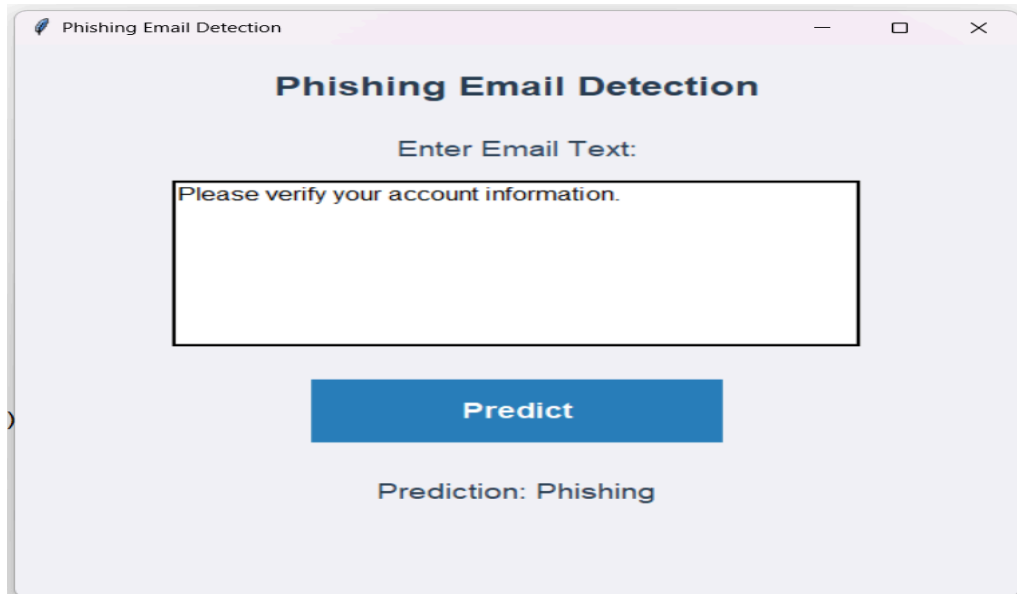
Phishing Email Detection

Enter Email Text:

Predict

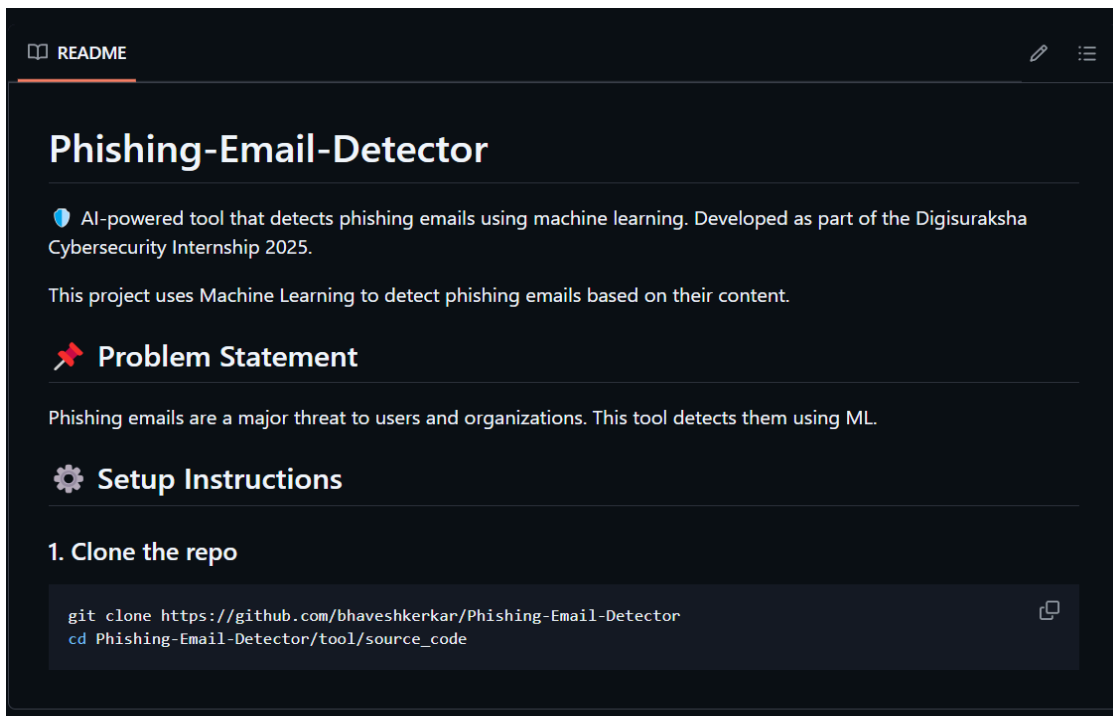
Prediction:

5. Classification Result –



The screenshot shows a web application window titled "Phishing Email Detection". The interface has a light purple background. At the top, the title "Phishing Email Detection" is displayed in a bold, dark font. Below the title, the text "Enter Email Text:" is shown. A large, empty text input field is positioned below this text. Inside the input field, the placeholder text "Please verify your account information." is visible. Below the input field, there is a prominent blue button with the word "Predict" in white. At the bottom of the interface, the text "Prediction: Phishing" is displayed.

6. GitHub Repository Page



The screenshot displays the GitHub repository page for "Phishing-Email-Detector". The page has a dark theme. At the top, there is a "README" tab. The main heading is "Phishing-Email-Detector". Below the heading, there is a description: "AI-powered tool that detects phishing emails using machine learning. Developed as part of the Digisuraksha Cybersecurity Internship 2025." followed by "This project uses Machine Learning to detect phishing emails based on their content." The page is divided into sections by horizontal lines. The first section is "Problem Statement" with a red star icon, containing the text "Phishing emails are a major threat to users and organizations. This tool detects them using ML." The second section is "Setup Instructions" with a gear icon. Under "Setup Instructions", there is a sub-section "1. Clone the repo" which contains a code block with the following commands:

```
git clone https://github.com/bhaveshkerkar/Phishing-Email-Detector
cd Phishing-Email-Detector/tool/source_code
```

11. References

1. Scikit-learn Documentation - <https://scikit-learn.org>
2. Kaggle Dataset - [Email Phishing Dataset](#)
3. TF-IDF Wikipedia - <https://en.wikipedia.org/wiki/Tf%E2%80%93idf>
4. Naive Bayes Classifier - <https://towardsdatascience.com>
5. NLP in Phishing Detection - ResearchGate Paper
6. Email Security Reports - Cisco Annual Report
7. NLTK Documentation - <https://www.nltk.org>
8. Google AI Blog – Fighting phishing with AI
9. IBM Security Blog – AI & Email Security
10. OWASP Foundation – Phishing Detection Standards

Useful Links

- ♦ **GitHub Repository:** [GitHub - AI-Powered Phishing Email Detector](#)
- ♦ **LinkedIn - Bhavesh Kerkar:** [LinkedIn Profile](#)
- ♦ **LinkedIn - Devraj Pujari:** [LinkedIn Profile](#)