IBM Innovation Center for Education

PHEMESOFT
Software That Matters
TRUSTED DELIVERY PARTNER for IBM ICE

RABBI ZIDNI 'ILMA

YENEPOYA
(DEEMED TO BE UNIVERSITY)
Recognised under Sec 3(A) of the UGC Act 1956
Accredited by NAAC with 'A' Grade

PHISHING EMAIL DETECTION SYSTEM

PROJECT SYNOPSIS
Phishing Email Detection Using Machine Learning

BACHELOR OF COMPUTER SCIENCE
CYBER FORENSIC, DATA ANALYTICS AND CYBER
SECURITY

SUBMITTED BY
Nuha Khatun
22BSCFDC32

GUIDED BY
Mr. Sumith Shukla

TITLE

Group Members:

1. Ayesha Hafseen
   Registration No.: 22BSCFDC10
   Mail Address: 22137@yenepoya.edu.in
   Course & Batch: BSC Cyber Forensic IBM, 2022-2025
2. Nuha Khatun
   Registration No.: 22BSCFDC32
   Mail Address: 23342@yenepoya.edu.in
   Course & Batch: BSC Cyber Forensic IBM, 2022-2025
3. Ishra Fathima
   Registration No.: 22BSCFDC19
   Mail Address: 23064@yenepoya.edu.in
   Course & Batch: BSC Cyber Forensic IBM, 2022-2025

INDEX

1. INTRODUCTION

In today's digital age, phishing has emerged as a significant and complex cybersecurity challenge. Phishing attacks involve malicious actors impersonating trusted entities to deceive individuals into revealing sensitive information like passwords, financial details, and login credentials. The consequences of falling victim to phishing are severe, including financial fraud, identity theft, and data breaches. The sheer volume of digital interactions creates a vast attack surface, making phishing a critical social engineering problem that preys on human trust and urgency.

Traditional detection methods, relying on human vigilance or static rule-based systems, are increasingly insufficient against the evolving sophistication of these attacks. Phishers continuously adapt their tactics, employing advanced social engineering techniques, novel obfuscation methods, and sophisticated technical subterfuge to bypass conventional security measures. This escalating threat necessitates an urgent need for advanced, proactive, and automated detection systems capable of keeping pace with the dynamic nature of cybercrime.

This project, the Phishing Email Detection System, directly addresses this imperative. It is an innovative solution meticulously engineered to mitigate the critical threat of phishing by leveraging the power of advanced machine learning (ML) techniques. Unlike rigid, rule-based systems that are easily circumvented by novel phishing variants, a machine learning approach offers inherent adaptability and the capacity to learn from vast amounts of data. This enables the system to identify subtle, complex patterns indicative of malicious intent that might elude human detection or traditional filters. The primary design objective is to classify emails as either phishing or legitimate in real-time, thereby serving as a crucial automated defense mechanism. By providing rapid and accurate classification, this system aims to significantly reduce the window of opportunity for attackers and enhance the overall resilience of digital communication channels. Its widespread adoption is not merely a technical advancement but a fundamental necessity to safeguard individuals and organizations from the devastating impacts of financial fraud, identity theft, and data breaches in an increasingly interconnected world.

This report provides a comprehensive overview of the Phishing Email Detection System. Section 2 details the methodology, covering data acquisition, preprocessing, feature engineering, model selection, training, evaluation, deployment, and user interface development. Section 3 enumerates the essential hardware and software facilities. Finally, Section 4 lists the references.

2. METHODOLOGY

The Phishing Email Detection System employs a systematic process to identify and classify phishing emails, leveraging machine learning principles. The methodology begins with defining the objective: to accurately classify emails as phishing or legitimate.

1. **Data Collection and Preparation:**
   - A sample dataset of over 50 email texts, labeled as 'phishing' or 'legitimate', was generated to simulate realistic email scenarios including common phishing tactics like fake login links, account warnings, and prize claims.
   - Data preprocessing involves cleaning and normalizing the text: lowercasing, replacing URLs with "url", email addresses with "email", and numbers with "number", and removing punctuation.
   - Labels are encoded numerically ('phishing' as 1, 'legitimate' as 0).
   - The dataset is split into 80% training and 20% testing sets for model development and unbiased evaluation.
2. **Feature Engineering:**
   - TF-IDF (Term Frequency-Inverse Document Frequency) vectorization is used to convert cleaned email texts into numerical feature vectors. This technique quantifies the importance of words, balancing their frequency in a document with their rarity across the entire dataset.
   - The `TfidfVectorizer` is configured to remove English stop words and limit features to the top 5000 terms, focusing on the most informative words.
3. **Machine Learning Model Selection and Training:**
   - Logistic Regression was chosen as the primary classification algorithm due to its interpretability, efficiency, and effectiveness in binary text classification.
   - The model is trained on the TF-IDF vectorized training data, learning patterns indicative of phishing or legitimate emails.
4. **Model Evaluation:**
   - The trained model's performance is evaluated on the unseen test data using metrics such as accuracy, precision, recall, and F1-score. Initial results showed 100% accuracy on the small test set, serving as a promising proof of concept.
5. **Model Deployment and Real-time Prediction:**
   - The trained Logistic Regression model and the fitted TF-IDF vectorizer are serialized using `joblib` for persistence, allowing them to be loaded for real-time predictions without retraining.
   - A `predict_email` function integrates preprocessing and model inference to classify new email content instantly.
6. **User Interface Development:**

- A user-friendly web interface is developed using the Gradio library, enabling users to easily input email content and receive immediate classification results.

In summary, this methodology provides a comprehensive approach to building a machine learning-based phishing email detection system, from data preparation and feature engineering to model training, evaluation, and user-friendly deployment.

3. FACILITIES REQUIRED

The development and operation of the Phishing Email Detection System require specific hardware and software facilities.

Tools:

- **Programming Language:** Python (version 3.x recommended).
- **Libraries & Frameworks:**
  - `Scikit-learn`: For machine learning algorithms (Logistic Regression), feature extraction (TF-IDF), and evaluation metrics.
  - `Pandas` and `NumPy`: For data processing and numerical operations.
  - `Gradio`: For developing the user-friendly web interface.
  - `Joblib`: For model serialization and persistence.
  - `re` (Regular Expressions) and `string`: For text preprocessing tasks.
- **Development Environment:** Integrated Development Environments (IDEs) such as VS Code, PyCharm, or Jupyter Notebooks.
- **Version Control:** Git (recommended for managing code changes and collaboration).

Hardware Requirements:

- **Computers:** Standard personal computers or workstations with sufficient CPU and RAM (at least 8GB) are adequate for data processing, model training, and running the Gradio interface.
- **Storage:** Adequate SSD/HDD space to store the dataset, trained models (`.pkl` files), and other project files.

4. REFERENCES

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Scikit-learn Documentation (scikit-learn.org)
- Gradio Documentation (gradio.app)
- Pandas Documentation (pandas.pydata.org)