JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SafeGuardProbe

[1]Parshva Chetan Doshi, [2]Darsh Bharat Patel, [3]Darsh Bhavesh Patel,

[4]Omkar Pravin Solanki, [5]Shwetambari Borade

[1]Student, [2]Student, [3]Student, [4]Student, [5]Teacher
[1]Cyber Security, [2]Cyber Security,[3]Cyber Security,[4]Cyber Security, [5]Cyber Security
[1]Shah And Anchor Kutchhi Engineering College, Mumbai, India

Abstract : In the midst of rising cyber attacks, it's crucial to protect private data to maintain integrity, confidentiality and availability. SafeGuardProbe is a solution designed to tackle phishing attacks, a major threat to data security. This tool uses advanced technologies like Python, Machine Learning, and Natural Language Processing (NLP). It checks if a given website link is safe or potentially phishing. SafeGuardProbe also analyzes emails using NLP. This means it can understand sentiments, recognize important information, and even figure out the main topics in emails. If it finds a phishing link or something suspicious in emails, it stores this information to respond quickly to new threats. SafeGuardProbe's impact is widespread. It not only warns users about potentially unsafe websites but also enhances email security by understanding and responding to email threats. This all-in-one approach significantly boosts cyber security, preventing sensitive data theft and strengthening defenses against various cyber risks.

IndexTerms **- Phishing, Ensemble learning, Email Text Analysis, Natural Language Processing (NLP)**

Introduction

In the ever-evolving landscape of technology, it is imperative to employ innovative strategies to safeguard both personal and organizational information. Introducing SafeGaurdProbe – a solution created to tackle the issues of data integrity and confidentiality. As the name suggests, SafeGaurdProbe specializes in detecting malicious intent, particularly combating the prevalent issue of phishing. Phishing is a type of malicious attack in which the attacker sends emails to the victim. These email contains malicious links to external websites which are designed to steal sensitive information from the victim. In a world where attackers disguise harmful links as legitimate, our model excels in assessing the credibility of mails. This cutting-edge technology serves as a robust defense against data breaches, preventing the theft of sensitive financial and personal information. SafeGaurdProbe becomes a valuable asset, enhancing overall security. With the surge in online threats and scams, especially with the expanding reach of the Internet, SafeGaurdProbe emerges as an effective tool in identifying and exposing hoax links, providing timely alerts to users.

I. Literature Survey & Technology Stack

Literature Survey

In this Literature Survey, twenty literature articles have been examined and compared. [1] This study suggests a way to classify non-spam messages according to language detection and sender ID or number. You can choose to translate or disregard the categorised messages. Natural Language Processing (NLP) approaches are used to analyse information for features including sentiment, spelling correction, and grammatical problems before generating replies in the user's preferred language. [2] The author of this paper presents ML-PE-TA, an efficient solution for analyzing email texts and detecting phishing. Achieving an impressive 96% accuracy, it utilizes a minimal set of essential features. The approach also employs dimensionality reduction, outperforming existing methods with higher feature counts and lower accuracy rates. The results highlight ML-PE-TA's efficacy in accurate phishing email analysis with reduced complexity. [3] This paper provides a user-friendly introduction to natural language processing (NLP), focusing on practical Python programming skills. Emphasizing tasks like predictive text, email filtering, summarization, and translation, the author covers accessing annotated datasets, using linguistic data structures, and employing key algorithms for content and structure analysis. With numerous examples and exercises, the paper guides readers in extracting information, analyzing linguistic structures, and integrating techniques from linguistics and AI. Utilizing Python and the Natural Language Toolkit (NLTK), it serves as a valuable resource for web development, multilingual news analysis, and linguistic documentation. [4] The author of this study highlights the ever-present and ever-changing threat posed by phishing, emphasising the effects it has on both people and brands. It offers a thorough analysis of the state of the art at the moment and emphasises how important it is to detect phishing attempts effectively. Detection methodologies are classified in the discussion into three categories: machine learning, list-based, and similarity-based techniques. The research highlights research gaps for more

investigation, evaluates used datasets, and examines suggested detection techniques within each category. [5] The author of this study highlights the ever-present and ever-changing threat posed by phishing, emphasising the effects it has on both people and brands. It offers a thorough analysis of the state of the art at the moment and emphasises how important it is to detect phishing attempts effectively. Detection methodologies are classified in the discussion into three categories: machine learning, list-based, and similarity-based techniques. The research highlights research gaps for more investigation, evaluates used datasets, and examines suggested detection techniques within each category. [6] This paper introduces Phish Responder, a solution to address the issue of phishing and spam vulnerability within organisations. It makes use of natural language processing and deep learning through a hybrid machine learning technique. With the LSTM model for text-based datasets and the MLP model for numerical-based datasets, the experiment shows exceptional average accuracy of 99% and 94%, respectively. The statistical superiority of the numerical-based method is highlighted by comparative assessments and an independent t-test, highlighting Phish Responder's efficiency in enhancing the detection of spam and phishing emails. [7] In order to combat the security risk posed by phishing assaults, this study presents a novel strategy that uses natural language processing to analyse text. It does semantic analysis of the natural language text with an emphasis on identifying offensive remarks inside the assault in order to determine malevolent intent. An important advancement in the field of phishing attack detection is made by evaluating the approach's efficacy on a benchmark set of phishing emails. [8] This paper shows a similar approach as but when it classifies a URL as phishing, it will store the list of the phishing websites in a txt file. [9] Its method for identifying phishing URLs is noteworthy. The proportionate distance between the input URL and the database URL is calculated. It is determined to be phishing if there is a significant discrepancy between the two. In addition, it makes use of the Favicon Images Recognition Algorithm and the False Positive and False Negative pre-existing approaches. [10] Used a simple but effective approach using Machine Learning to detect Phishing URLs.

[11] Uses 5 different Machine Learning Algorithms to Classify the URLs. Presence of 2 different datasets and Random Forest produced the highest accuracy of 95% and 96.8% on both the datasets. [12] discusses univariate feature selection methods such as ANOVA, Chi-square, and IG. enabling the use of univariate statistical tests to choose the optimal characteristics. This makes it possible to use KNN to attain an astounding accuracy of 99.8%. [13] This paper extracts several features from the URL to Classify it as Phishing or Legitimate. But after the recognition is completed, a snapshot of the web page is extracted and is compared with the regular web page snapshot to implement the recommendation of the original regular web page of the phishing web page. [14] This paper provides a URL Based as well as a Network Based Solution to detect Phishing URLs. A total of 10 Machine Learning Algorithms are used and their results are compared with each other. [15] Tokenizer and Count Vectorizer aggregate words are used in the model because some words in URLs—like "virus," ".exe," and so forth—are more important than others. The software determines a text's tag, like an email or news item, using the Bayes theorem. [16] This study clarifies the key elements that have been shown to be reliable and successful in identifying phishing websites. The URL features are divided into three groups: features based on the address bar, features based on abnormalities, and features based on HTML and JavaScript. The URLs are filtered using a total of thirty parameters, which divides them into three categories: 0 is suspicious, 1 is legitimate, and -1 is phishing.

[17] In this paper, a Random Forest model was employed to detect phishing websites with an impressive 99.55 accuracy. A technique known as 'Principal Component Analysis Random Forest' was utilised, outnumbering numerous other machine learning techniques and providing greater accuracy. This study suggested machine learning characteristics that included wrapper features. [18] The research provides a machine learning model based on gradient boosting approaches that is lightweight in nature. It is made up of a new dataset that concentrates solely on URL-based attributes, which limits its operations. [19] The following paper presented the 'Ensemble Learning' methodology for web frameworks. A dataset comprised of data from PhishTank, Alexa, and Kaggle. They employed SMOTE analysis (Synthetic Minority Oversampling Technique) to enhance the number of cases in the dataset. Machine Learning techniques such as K-means, Random Forest, Decision Tree, CatBoost classifier, LightGBM classifier, AdaBoost, and voting classifier were tried, with CatBoost achieving the greatest accuracy of any of them. [20] The model's major purpose is to focus on social engineering attacks, including phishing. They used machine learning to compare various strategies. They comprised Decision Tree (DT), Random Forest (RF), XGBoost, Multilayer Perceptrons, K-Nearest Neighbours, Naive Bayes, AdaBoost, and Gradient Boosting, with XGBoost achieving the greatest accuracy of 96.6. The dataset included emails, links, URLs, and personal information.

Technology Stack

This project utilizes technology, such as:

1. Python
2. Git

The entire project is coded in Python, and everything has been built from scratch, including implementation of Machine Learning, Frontend, and Backend components. We have also utilized Git version control for collaborative development, ensuring effective management of the project's codebase and version history.

II. RESEARCH METHODOLOGY

Natural Language Processing (NLP) plays a pivotal role in email text analysis. By leveraging advanced algorithms, NLP enables computers to comprehend and extract meaningful insights from the unstructured nature of email content.

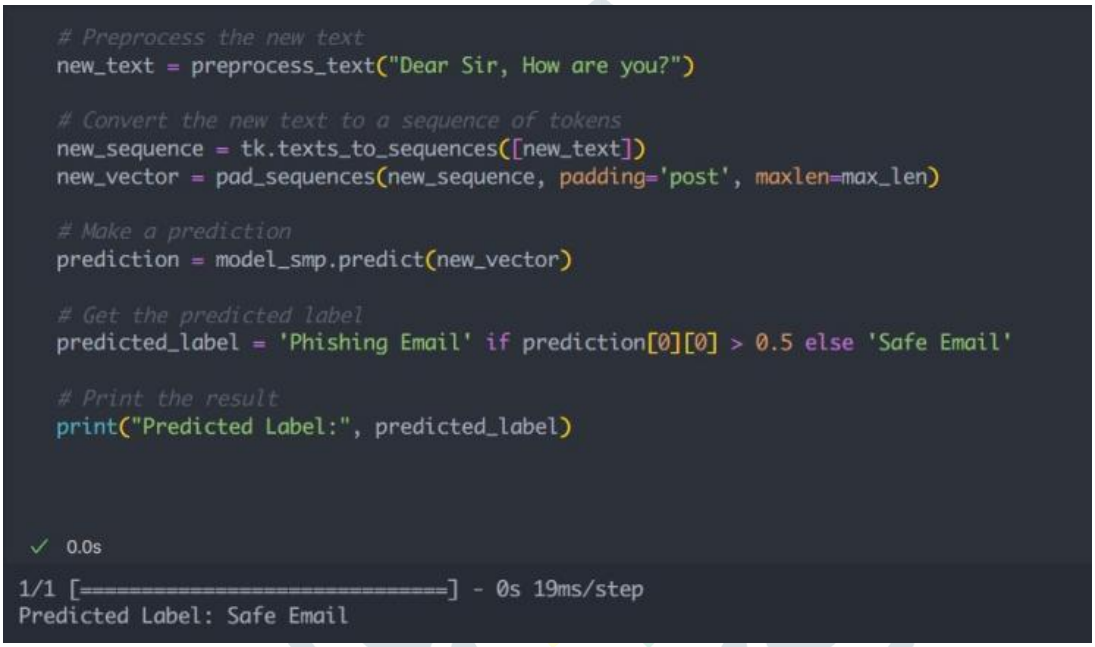
In the context of email text analysis, NLP can be applied for various purposes such as for Sentiment Analysis where in NLP algorithms can assess the sentiment expressed in emails, helping to gauge the emotional tone of the communication. Named Entity Recognition (NER) where NLP can identify and categorize entities mentioned in emails, such as names, locations, organizations, and more, providing valuable contextual information. Topic Modeling in which NLP techniques can be used to

identify the main themes or topics within email conversations, aiding in understanding the focus of communication. Language Translation in which NLP algorithms can support language translation, enabling communication across language barriers in global email interactions. Email Categorization where in NLP can categorize emails into different groups based on their content, facilitating efficient email organization and management. Keyphrase Extraction where Identifying key phrases within emails helps in summarizing and understanding the most important points of the communication.

By incorporating NLP into email text analysis, we can derive actionable insights from the vast amount of information exchanged through email communications.

We made use of Recurrent Neural Network(RNN) for Natural Language Processing(NLP). RNNs are a sort of neural network built for sequential data, making them ideal for jobs requiring the order of input data, such as natural language processing.
It includes loops to allow information to persist over time steps, allowing them to capture dependencies in sequential data. The fundamental unit of an RNN is a neuron with a self-connected recurrent edge. This loop transfers information from one step in the sequence to the next.

III. RESULTS AND DISCUSSION



The above image shows the implementation of the model. At the bottom of the image, we can see the result of email in predicted label as safe email meaning the email received is not a phishing email.
We successfully developed a phishing detection model using Simple RNN for prediction, achieving an impressive accuracy of 92.76%. The model's efficacy in distinguishing between phishing and legitimate mail underscores its robust performance. In addition, the integration of regex for data cleaning contributed to refining the input data, enhancing the model's ability to make accurate predictions. This accomplishment signifies a significant step forward in creating an effective and reliable phishing detection system, with the combined use of machine learning and data preprocessing techniques yielding promising results.

Figures and Tables

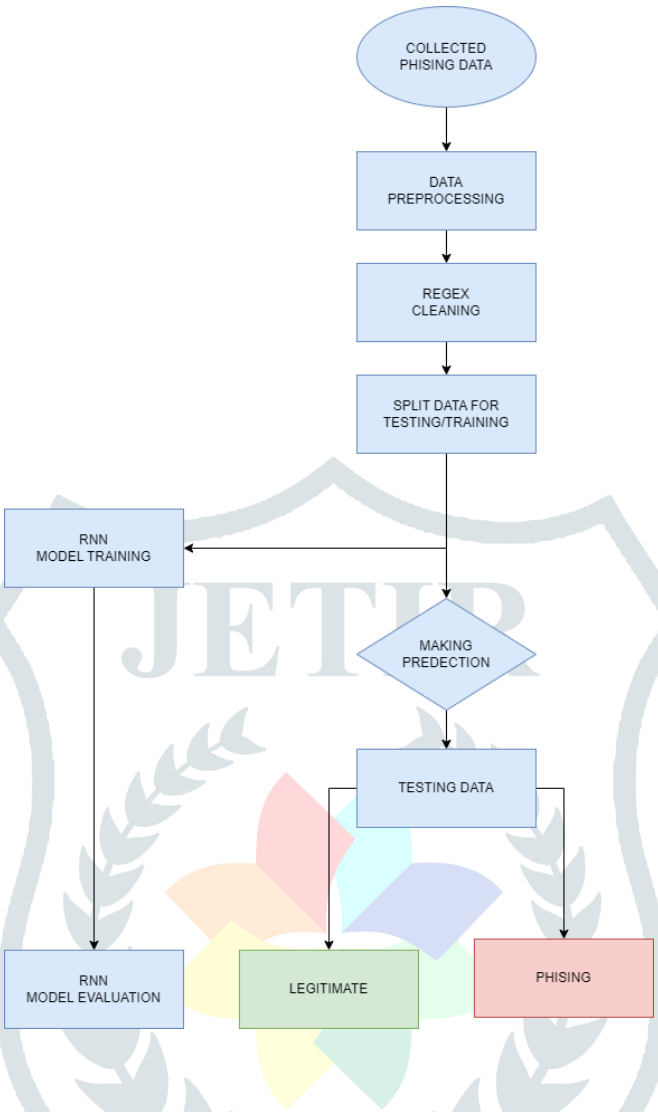


Table 1 Accuracy Table

Sr. no	Classifier	Accuracy
1	Simple RNN(Recurrent Neural Network)	92.76%

IV. Conclusion and Future Work

In summary, SafeguardProbe signifies a significant advancement in cybersecurity, presenting an efficient system for detecting phishing in mail. By using RNN for prediction SafeguardProbe enables users to quickly assess the legitimacy emails, providing protection against online threats. The project has demonstrated its effectiveness in bolstering online security and assisting users in making informed decisions during web navigation.

Looking forward, SafeguardProbe anticipates promising opportunities. Plans include expanding its reach through the development of browser extensions and mobile applications, ensuring users have on-the-go protection. Furthermore, ongoing enhancements will incorporate new features to adapt to emerging phishing techniques and digital threats. Through this dedication to innovation, SafeguardProbe aims to stay at the forefront of the fight against online malice, fostering a safer and more secure digital environment for everyone.

References

[1] C. Sathish, A. Mahesh, N. S. Karpagam, R. Vasugi, J. Indumathi and T. Kanchana, "Intelligent Email Automation Analysis Driving through Natural Language Processing (NLP)," 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2023, pp. 1612-1616, doi: 10.1109/ICEARS56392.2023.10085351.

[2] S. M. M. Ahammad, T. Raviteja, J. Koushik, P. V. Dinesh and A. Ashok, "Machine Learning Approach Based Phishing Email Text Analysis (ML-PE-TA)," 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), Kannur, India, 2022, pp. 1087-1092, doi: 10.1109/ICICICT54557.2022.9917765.

[3] E. Benavides, W. Fuertes, S. Sánchez-Gordón, D. Nuñez-Agurto, and G. Rodríguez-Galán, "A Phishing-Attack-Detection Model Using Natural Language Processing and Deep Learning," Apr. 2023, doi: 10.3390/app13095275.

[4] R. Zieni, L. Massari, and M. C. Calzarossa, “Phishing or Not Phishing? A Survey on the Detection of Phishing Websites,” Jan. 2023, doi: 10.1109/access.2023.3247135.

[5] S. A. Salloum, T. Gaber, S. Vadera, and K. Shaalan, “A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques,” Jan. 2022, doi: 10.1109/access.2022.3183083.

[6] M. Dewis and T. Viana, “Phish Responder: A Hybrid Machine Learning Approach to Detect Phishing and Spam Emails,” Jul. 2022, doi: 10.3390/asi5040073.

[7] T. Peng, I. G. Harris, and Y. Sawa, “Detecting Phishing Attacks Using Natural Language Processing and Machine Learning,” Jan. 2018, doi: 10.1109/icsc.2018.00056.

[8] Helmi, Rabab & Md Johar, Md Gapar & Hafiz, Muhammad. (2023). Online Phishing Detection Using Machine Learning. 1-4. 10.1109/ICAISC56366.2023.10085377.

[9] M. Mohammed, K. K. Prasanth and S. V. Sai Subhash, "Phishing Detection Using Machine Learning Algorithms," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2022, pp. 921-924, doi: 10.1109/ICSSIT53264.2022.9716269.

[10] M. Shoaib and M. S. Umar, "URL based Phishing Detection using Machine Learning," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2023, pp. 1-7, doi: 10.1109/ISCON57294.2023.10112184.

[11] M. Aljabri and S. Mirza, "Phishing Attacks Detection using Machine Learning and Deep Learning Models," 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA), Riyadh, Saudi Arabia, 2022, pp. 175-180, doi: 10.1109/CDMA54072.2022.00034.

[12] S. Mohanty, M. Sahoo and A. A. Acharya, "Predicting Phishing URL Using Filter based Univariate Feature Selection Technique," 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2022, pp. 1-5, doi: 10.1109/ICCSEA54677.2022.9936298.

[13] W. Bai, "Phishing Website Detection Based on Machine Learning Algorithm," 2020 International Conference on Computing and Data Science (CDS), Stanford, CA, USA, 2020, pp. 293-298, doi: 10.1109/CDS49703.2020.00064.

[14] A. Ghimire, A. Kumar Jha, S. Thapa, S. Mishra and A. Mani Jha, "Machine Learning Approach Based on Hybrid Features for Detection of Phishing URLs," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 954-959, doi: 10.1109/Confluence51648.2021.9377113.

[15] G. K. Kamalam, P. Suresh, R. Nivash, A. Ramya and G. Raviprasath, "Detection of Phishing Websites Using Machine Learning," 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2022, pp. 1-4, doi: 10.1109/ICCCI54379.2022.9740763.

[16] R. M. Mohammad, F. Thabtah and L. McCluskey, "An assessment of features related to phishing websites using an automated technique," 2012 International Conference for Internet Technology and Secured Transactions, London, UK, 2012, pp. 492-497.

[17] A. Odeh, I. Keshta and E. Abdelfattah, "Machine LearningTechniquesfor Detection of Website Phishing: A Review for Promises and Challenges," 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), NV, USA, 2021, pp. 0813-0818, doi: 10.1109/CCWC51732.2021.9375997.

[18] T. A and A. John, "Phishing Website Detection Using LGBM Classifier With URL-Based Lexical Features," 2022 IEEE Silchar Subsection Conference (SILCON), Silchar, India, 2022, pp. 1-7, doi: 10.1109/SILCON55242.2022.10028793.

[19] N. Puri, P. Saggar, A. Kaur and P. Garg, "Application of ensemble Machine Learning models for phishing detection on web networks," 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2022, pp. 296-303, doi: 10.1109/CCiCT56684.2022.00062.

[20] S. Alrefaai, G. Özdemir and A. Mohamed, "Detecting Phishing Websites Using Machine Learning," 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 2022, pp. 1-6, doi: 10.1109/HORA55278.2022.9799917.