International Journal of
Technology and Systems
(IJTS)

Combating Phishing in Kenya: A Supervised Learning Model for Enhanced Email Security in Kenyan Financial Institutions

Asiema Mwavali

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 4, No.3, pp 23 – 36, 2024

www.iprjb.org

Combating Phishing in Kenya: A Supervised Learning Model for Enhanced Email Security in Kenyan Financial Institutions

Asiema Mwavali
Department of Computing and Informatics, Technical University of Kenya, Nairobi

Article History

Received 16th May 2024

Received in Revised Form 20th June 2024

Accepted 30th July 2024

How to cite in APA format:

Mwavali, A. (2024). Combating Phishing in Kenya: A Supervised Learning Model for Enhanced Email Security in Kenyan Financial Institutions. *International Journal of Technology and Systems*, *9*(4), 23–36. https://doi.org/10.47604/ijts.2820

Abstract

Purpose: Phishing is a serious cybercrime problem that puts people and organizations at risk all around the world, especially in Republic of Kenya's financial institutions. Modern solutions are required because traditional security measures are being challenged by the growing sophistication of phishing attempts. The goal of this thesis is to use artificial intelligence (AI) to create a supervised machine learning model that will alleviate phishing email attacks in the Kenyan financial institutions. Attackers frequently use phishing emails as a means of obtaining unauthorized access to private information, including login credentials, financial information, and personal information. This can lead to identity theft, reputational harm, and monetary losses.

Methodology: The suggested framework focuses on using supervised machine learning techniques to recognize and stop phishing emails with accuracy. Four primary parts make up the framework: response, email filtering, feature extraction, and categorization. Four primary steps are involved in developing the framework: gathering data, preparing the data, training the model, and deployment. A thorough dataset of phishing and non-phishing emails is compiled throughout the data collecting phase from a variety of sources, including as public databases, forums on the dark web, and emails from financial institutions staff. Cleaning, classifying, and preprocessing the gathered data are all part of the data preprocessing step, which makes sure the data is appropriate for training the model. Supervised machine learning methods are used in the model training phase to create a reliable detection model.

Findings: The system is tested against a dataset of phishing and non-phishing emails particular to the Kenyan financial sectors. The framework's effectiveness is assessed using performance indicators such as accuracy, precision, recall, and the F1-score. The results reveal that the system can correctly detect new phishing emails that were not previously included in the training dataset, demonstrating its adaptability to emerging threats.

Unique Contribution to Theory, Practice and Policy: By offering an effective mechanism for detecting and alleviating phishing email attacks, the proposed framework considerably minimizes the risk of data breaches and financial losses in the banking sector.

Keywords*:* *Supervised Machine Learning, Model, Alleviating, Phishing Email, Training Dataset, Financial Institutions, Machine Learning Algorithm, Email Filtering, Feature Extraction, Classification, Accuracy, Precision, Recall, F1-score.*

JEL Codes: *G21, D83, L86*

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 4, No.3, pp 23 – 36, 2024
www.iprjb.org

INTRODUCTION

Phishing emails attacks in the financial institutions, including banking industry in Kenya are carried out through various methods, including email, SMS, and phone calls. Email remains the most common method, and attackers usually send emails that appear to come from legitimate sources such as banks, financial institutions, or government agencies. Once a system user opens the attachment, the malware infects the device, allowing the attackers to steal sensitive information or take control of the device.

In Kenya, Phishing email attacks are becoming increasingly sophisticated, making it difficult for individuals and organizations to detect and prevent them. One common method is spear-phishing, where attackers target specific individuals within the bank using personalized messages that appear to be from a trusted source, such as a colleague or superior. Another technique used in phishing attacks is social engineering, where assailants exploit human psychology to trick people into revealing subtle info or executing activities that could compromise the safety of the financial institutions systems. Some of the websites may clearly look fake with;

- Suspicious or mismatched sender email address
- Unexpected or urgent message subject
- Generic or non-personalized greeting
- Request for personal information, such as account numbers, passwords
- Links or attachments that are not expected or from an unknown source
- Poor grammar, misspelling, or formatting errors
- Threats or warnings of consequences if action is not taken

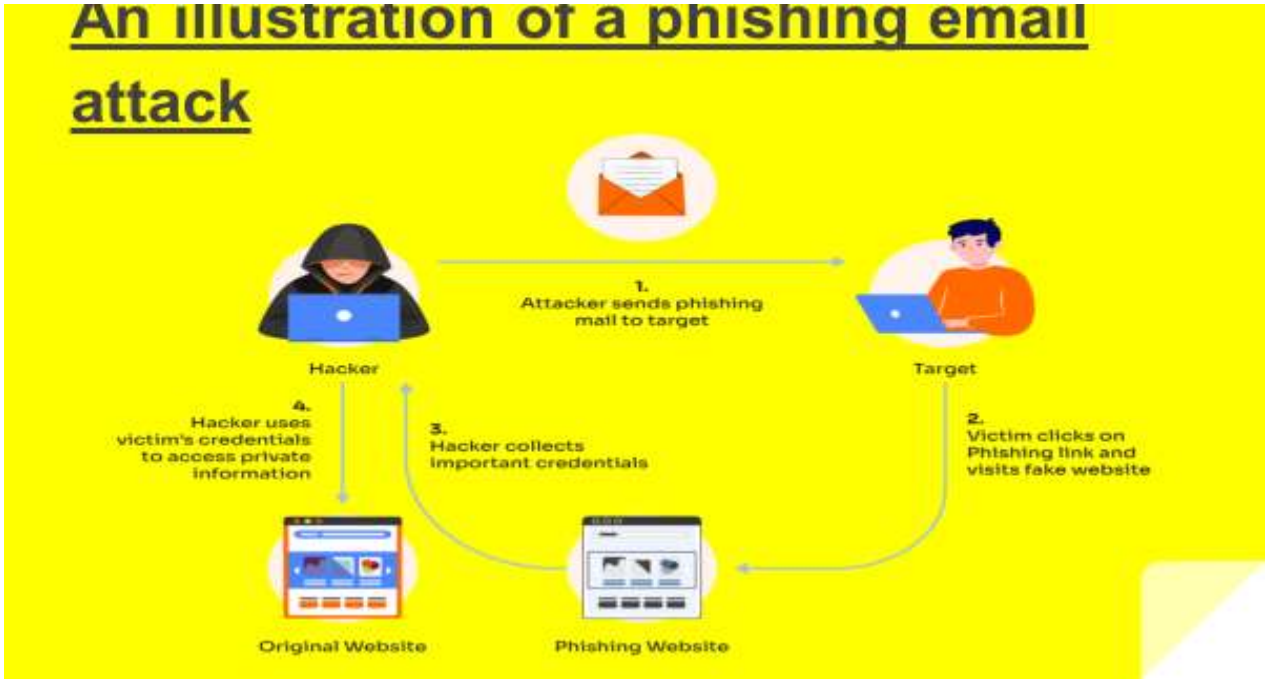


Figure 1: Illustration Showing a Phishing Email Attack

Supervised Learning

This thesis will utilize a supervised learning approach for phishing email detection. In supervised learning, algorithms learn from a dataset where each data point (email) is paired with a corresponding label (phishing or legitimate). By analyzing these labeled examples, the model learns to identify patterns that differentiate between the two categories. This allows it to classify new, unseen emails with a certain degree of accuracy.

Challenges of Labeled Data Acquisition

While supervised learning offers advantages for phishing email detection, one of its key challenges lies in acquiring a large, high-quality dataset of labeled emails. Accurately labeling emails as phishing or legitimate requires expertise and can be time-consuming. Phishing techniques are constantly evolving, making it crucial to have a dataset that reflects these changes.

Strategies for Overcoming Challenges

Several strategies can be employed to address the challenge of labeled data:

Collaboration with Security Experts: Partnering with cybersecurity professionals can provide access to real-world phishing email samples already classified.

Crowdsourcing: Utilizing crowdsourcing platforms can help gather labeled data, although quality control measures are essential.

Synthetic Data Generation: Techniques like generating synthetic phishing emails based on real-world examples can help augment the dataset.

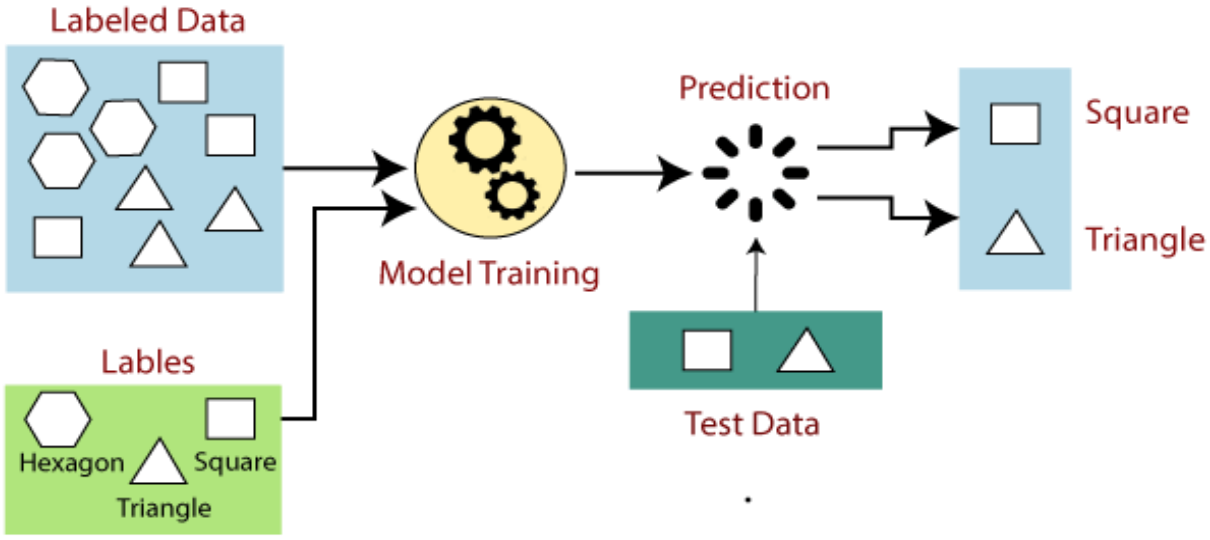


Figure 2: Illustration Showing Supervised Machine Learning Techniques

Bunching Algorithm

Bunching Algorithm will be utilized for this chore. Bunching is a machine-learning method that comprises assemblage comparable data points into clusters founded on their likenesses or alterations. The aim of the bunching is to recognize patterns and group them based on their structures (like the links, subject line, any attachments, and the body of the email) of which might not be instantly seeming to human viewers.

Clustering: Refer to clustering algorithm to the preprocessed data set of emails. One general algorithm that can be utilized for this resolution is the random forest model's algorithm. The amount of clusters can be decided by using methods like an elbow technique or silhouette score.

Cluster Scrutiny: Scrutinize resultant clusters to categorize resemblances as well as alterations among the emails. This can be carried out by envisaging clusters utilizing procedures like t-SNE or PCA. The clusters can also be physically examined to recognize any patterns or physiognomies that are revealing of phishing bouts.

Statement of the Problem

Phishing emails pose a significant threat to the Kenyan financial institutions, with a notable rise in both frequency and sophistication in recent years. These attacks have led to substantial financial losses for financial sectors and their clients. Existing phishing email alleviation frameworks and models exhibit several vulnerabilities that include;

Inability to Detect Emergent Bouts, **High False Positive Rates** and Lack of Customization for Kenyan financial institutions recent surveys such as the **PwC Kenya Economic Crime Survey 2023** and **KPMG Kenya Fraud and Risk Survey 2022** highlight the prevalence of email phishing as a leading economic crime and fraud risk in Kenya's business landscape. These findings underscore the urgent need for robust, context-aware solutions.

Artificial intelligence (AI) offers promising avenues to address these challenges. AI can empower the development of adaptive models capable of learning from and responding to new phishing tactics.

Solution

This thesis proposes a supervised learning approach to develop an AI-powered phishing email alleviation framework for the Republic of Kenyan Financial Institutions.

Supervised learning encompasses training a model using labeled data, where the model learns patterns from known examples to make predictions on novel, hidden data.

This supervised learning approach harnesses AI to overcome the limitations of existing frameworks, offering a tailored solution to enhance the security posture of Kenyan banks against phishing attacks.

By leveraging labeled data and advanced machine learning techniques, this framework aims to provide proactive defense measures against evolving cyber threats in the banking sector.

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 4, No.3, pp 23 – 36, 2024
www.iprjb.org

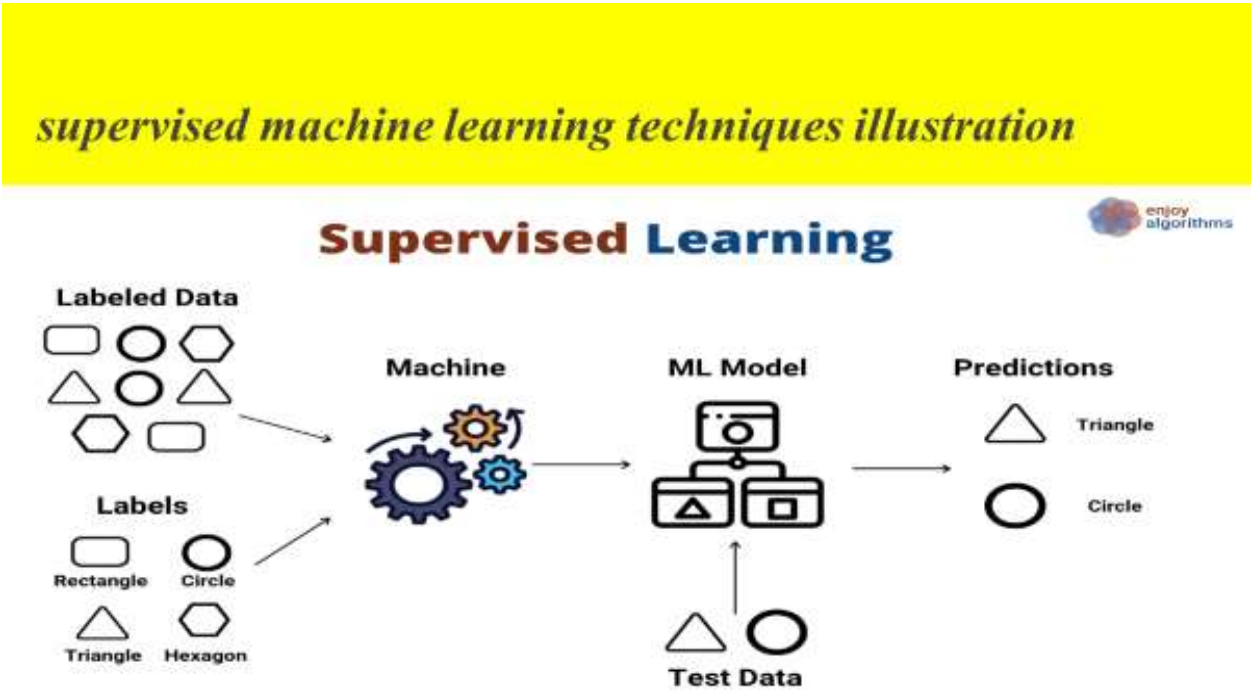


Figure 3: Illustration Showing Supervised Machine Learning Techniques

AI-powered solutions can also provide an additional layer of protection to customers. For example, financial institutions like banks can use AI-powered chatbots to interact with customers and detect potential phishing attempts. These chatbots can use natural language processing (NLP) algorithms to analyze the customer's messages and detect any suspicious behavior or language patterns. If the chatbot identifies a potential phishing attempt, it can warn the customer or direct them to a human representative for further assistance.

Figure 4: AI-Powered Chatbots Respond Intelligently to User Inquiries

LITERATURE REVIEW

Recent Cases of Phishing Attacks

Phishing attacks continue to pose a substantial threat to Kenya's financial institutions, resulting in data breaches and financial losses. This section investigates current phishing attempts on Kenyan Financial institutions and illustrates the evolving strategies used by perpetrators.

A surge in phishing attempts: The Communications Authority of Kenya (CA) reported in February 2024 that over 100 million cyber threats were discovered in Kenya during a three-month period (October-December 2023). This represents a 943 percent increase over the previous quarter, illustrating the worrying surge in cyber dangers, particularly phishing efforts.

Targeted attacks: Hack on e-Citizen Platform (June 2023): A recent hack targeted Kenya's e-Citizen platform, a critical government gateway for accessing services such as passport applications and driver's licenses. This highlights the increasing complexity of attackers targeting vital infrastructure to interrupt essential services and perhaps steal sensitive data. (BBC News)

Nigerian Mobile Money Users Targeted by Phishing SMS (2021)

In 2021, Nigerian mobile money users were targeted by a widespread phishing SMS campaign. The attackers sent messages disguised as notifications from mobile money operators, prompting users to click on links that redirected them to fake websites designed to steal their login credentials and account balances. This incident underscores the growing threat of mobile-based phishing attacks, particularly in regions with high mobile phone penetration.

University of Chicago

Spear phishing attempts are harder to identify than conventional phishing assaults because they are more focused and customized. In order to carry out these assaults, the victim's personal information is researched, and emails purporting to be from a bank or other reliable source are created.

In July 2023, phishing emails purporting to be from Microsoft Office 365 support targeted University of Chicago in an effort to get login credentials. See https://security.uchicago.edu/phishing/

FACC

Security News, 2022 report indicates that in the year 2016 January a counterfeit memo from CEO of Austrian organization FACC, a worker fell for a phishing email bout (Reuters, 2016). It has been reported that an employee of the said company received an email message requesting for a transfer of more than 40million euros to a different account as share of "purchase project".

Recent Studies on using Artificial Intelligence

Recent studies have shown the effectiveness of using AI to mitigate phishing attacks. One study conducted by researchers at the University of Texas at San Antonio found that a machine learning algorithm was able to identify phishing emails with an accuracy of 94.3%. Another study conducted by researchers at the University of Portsmouth in the UK found that AI-based phishing detection systems were able to detect phishing attacks with a 99.9% accuracy rate.

A study by the University of Nairobi in 2019 examined the effectiveness of various anti-phishing techniques, such as two-factor authentication and user education.

In their study, Kabiru et al. (2021) developed an AI-based system for detecting phishing attacks in the banking sector. The authors used machine learning algorithms to analyze email headers and content and to identify phishing emails.

A study by the Journal of Information Security and Applications showed that the use of machine learning algorithms was effective in detecting phishing emails, with an accuracy rate of 97.2%.

In another study by the International Journal of Advanced Computer Science and Applications, the use of AI-based phishing detection techniques was found to be effective in reducing the success rate of phishing attacks. The study showed that the use of AI-based techniques reduced the success rate of phishing attacks from 40% to 2.5%.

Further studies by Hamdoun et al. (2021) proposed a deep learning-based framework for detecting phishing emails in the banking industry. The framework used a convolutional neural network (CNN) to extract features from the email content, and a recurrent neural network (RNN) to classify the email as legitimate or phishing. The study achieved an accuracy of 99.5% in detecting phishing emails.

Arul and Chandra (2021) proposed a machine learning-based framework for detecting phishing attacks in the banking industry. The framework used a combination of random forest and logistic regression algorithms to analyze various features of the email, such as sender address, subject line, and content. The study achieved an accuracy of 96.7% in detecting phishing emails.

Kilonzi et al. (2019) study "An Investigation of Phishing Attacks in the Banking industry in Kenya" found that existing anti-phishing frameworks were limited in their ability to detect and prevent new and sophisticated phishing attacks.

Mutinda and Mugambi (2020) in their study “Phishing Attacks in the Banking Sector in Kenya: An Analysis of the Current Situation" identified the lack of integration between different anti-phishing technologies as a major limitation of existing frameworks.

Okeyo and Mwongera (2020) study titled "Challenges and Opportunities of Email Phishing Attacks in the Banking industry in Kenya" identified several limitations of existing anti-phishing frameworks in Kenya. The study found that existing frameworks relied heavily on user education and awareness, which is often ineffective as users are not always able to distinguish between legitimate and phishing emails.

Munaiah et al. (2021) proposed a machine learning-based framework for detecting fraud in online banking transactions, while a study by Banerjee et al. (2020) proposed a deep learning-based framework for identifying malware in banking systems.

In Kenyan context, several banks have already begun to adopt AI-powered anti-phishing solutions. For example, Standard Chartered Bank Kenya has implemented an AI-powered anti-phishing system that uses machine learning algorithms to analyze user behavior and detect potential phishing attacks. Similarly, KCB Bank Kenya has implemented a multi-factor authentication system that uses AI to detect and prevent phishing attacks.

Research Gaps and Future Studies: Limited Dataset Diversity
Current Analysis: The study employs a dataset that includes phishing and non-phishing emails from public databases and bank personnel.
Research Gap: The dataset may not fully capture the wide variety of phishing strategies and linguistic variants used in phishing attempts, particularly those unique to distinct areas or cultural contexts.
Future study should focus on increasing the dataset to include a broader range of phishing samples, such as emails from different languages, regions, and demographic groupings. This would improve the model's robustness and generalizability across different contexts.

METHODOLOGY

The research methodology focuses on developing an AI-powered framework to mitigate phishing email threats in the Kenyan financial institutions. Conducted a broad review of current literature on phishing email bouts and AI-founded mitigation frameworks in Kenya. Identified gaps and relevant studies such as the "Kenya Banks Association (KBA) Cyber Security Report 2022" and "Central Bank of Kenya (CBK) Fraud Report 2021".

Area of Study: Concentrated research in Nairobi County due to budget constraints. Involved selected banks, employees

Data Collection

1. Gathered data from publicly available datasets, phishing emails from Kenyan financial institutions that included banks, and surveys/interviews with financial institutions clients and cybersecurity specialists.
2. Ensured quality through pilot testing, informed consent, and qualitative analysis tools.

Model Choice

- Selected supervised learning due to availability of labeled data and task classification requirements.
- Chose a Random Forest classifier for its effectiveness in email classification task

Model Training

- Preprocessed email data by cleaning, feature engineering, and vectorization.
- Trained the model, tuned hyperparameters, and evaluated performance metrics like accuracy and F1 score.

Model Development and Evaluation

- Developed AI-based phishing email classification models specific to Kenyan financial institution’s needs.
- Evaluated framework performance in detecting and blocking phishing emails, reducing false positives, and usability.

Ethical Considerations

- Upheld ethical standards throughout data collection, processing, and reporting.
- Ensured participant confidentiality, privacy, and informed consent were maintained.

DATA ANALYSIS AND RESULTS

Data preprocessing and feature engineering are vital stages in making the data for analysis and training the AI model successfully.

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 4, No.3, pp 23 – 36, 2024
www.iprjb.org

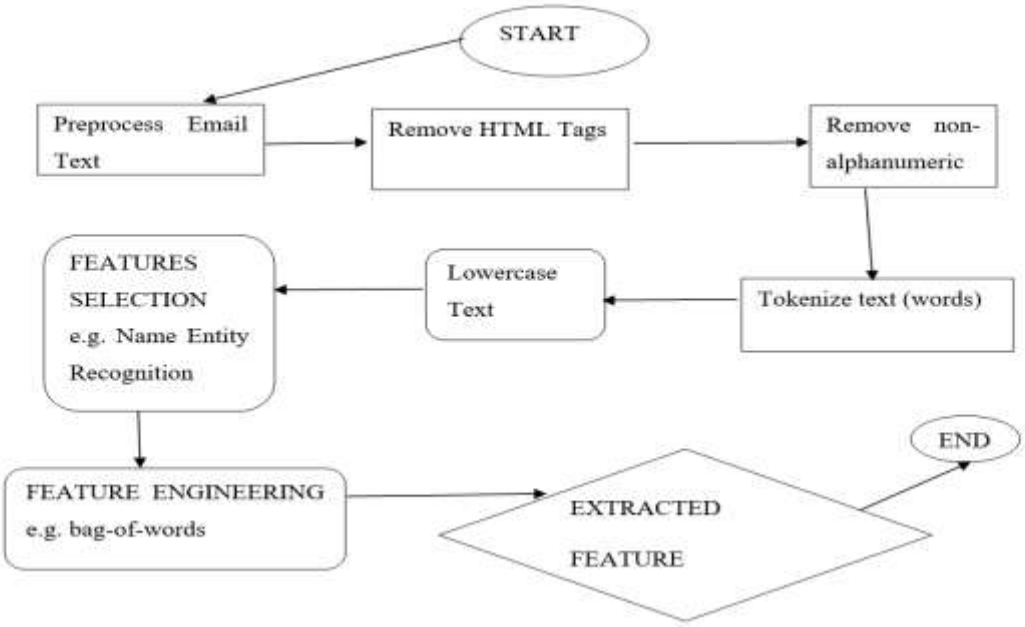


Figure 5: Shows Data Preprocessing and Feature Engineering

Data Preprocessing and Feature Engineering

In this chapter, we discuss the preprocessing steps and feature engineering techniques used to prepare our dataset for training the phishing email detection model. The raw data consists of email content labeled as either phishing (1) or legitimate (0).

Email Content	Label
As a valued KCB customer, you're invited to a special event on financial planning. RSVP today! Hello!	0
Urgent: Verify your account details to avoid suspension (KCB). Please respond ASAP!	1
Important: Your KCB account information needs updating to comply with new security regulations.	1
Get early access to new features with the latest update for your loan (KCB). Please respond ASAP!	0
Here are some tips to keep your KCB online banking safe from fraud and scams. Good morning!	0
Your KCB account balance is low. Top up your account for easy access to your funds. Hello!	0
As a valued KCB customer, you're invited to a special event on financial planning. RSVP today! Don't miss ou	0
As a valued KCB customer, you're invited to a special event on financial planning. RSVP today! Good mornin	0

Figure 6: Indicating Data Preprocessing and Feature Engineering

Data Cleaning and Transformation

To clean and transform the email data, we implemented the following steps:

Lowercasing: Converted all text to lowercase to ensure uniformity.

HTML Tag Removal: Removed any HTML tags that might be present in the email content.

Non-alphabetic Character Removal: Eliminated all non-alphabetic characters to focus on the words.

Tokenization: Split the text into individual words.

Stopword Removal: Removed common English stopwords that do not contribute to the meaning of the text.

Feature Selection and Extraction

Feature assortment includes recognizing the most pertinent features from the data that contribute to the forecast of phishing emails. This aids to enhance the model's performance and lessen computational problem.

We used a bag-of-words model to change the text data into numerical features that can be fed into the machine learning model. The CountVectorizer from sklearn was utilized for this purpose

Description of the AI Model

A Random Forest classifier was chosen for this task due to its robustness and effectiveness in handling text classification problems. The model was trained using the preprocessed and vectorized email data.

Training and Validation Data Split

The dataset was riven into training and testing sets with a 70-30 ratio, ensuring that both sets are stratified to maintain the original distribution of classes.

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 4, No.3, pp 23 – 36, 2024
www.iprjb.org

```
data = {
    'email_content': [
        'Urgent: Update your bank account details now!',
        'Your monthly statement is ready for viewing',
        'Win a free iPhone! Click here!',
        'Payment received: KES 5000 credited to your account',
        'Security alert: Unusual activity detected on your account',
        "Congratulations! You've won a cash prize. Click to claim",
        'Your loan application has been approved',
        'Important: Verify your account to avoid suspension',
        'New features available in your mobile banking app',
        'Limited time offer: Increase your credit limit today!',
        'Reminder: Your credit card payment is due soon',
        'Exclusive investment opportunity. Act now!',
        'Your account has been temporarily locked. Verify now',
        'Welcome to our new secure banking portal',
        'Fraud alert: Unauthorized transaction attempt'
    ],
    'label': [1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0]  # 1 for phishing, 0 for legitimate
```

Testing

```
test_emails = [
    "Urgent: Your account will be closed. Click here to prevent this!",
    "Your monthly bank statement is ready for download.",
    "You've won a prize! Click here to claim it now!",
    "Important notice: Update your account information immediately.",
    "Thank you for your recent transaction. Here's your receipt."
]
```

Model Training Process

After preprocessing and vectorizing the data, the Random Forest classifier was trained on the training set. The classifier was then evaluated using the test set, and its performance metrics were calculated.

Evaluation Metrics

The execution of the model was assessed using the following metrics:

Classification Report: Provides precision, recall, and F1-score for both classes.

Confusion Matrix: Visualizes the execution of the model by displaying the true positives, false positives, true negatives, and false negatives.

Visualization and Interpretation

Visualization of Vital Features and their correlation to phishing emails

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 4, No.3, pp 23 – 36, 2024
www.iprjb.org

To gain insights into the data and the model, several visualizations were created:

Class Distribution: Shows the balance between phishing and legitimate emails in the dataset.

Email Length Distribution: Compares the length of phishing and legitimate emails.

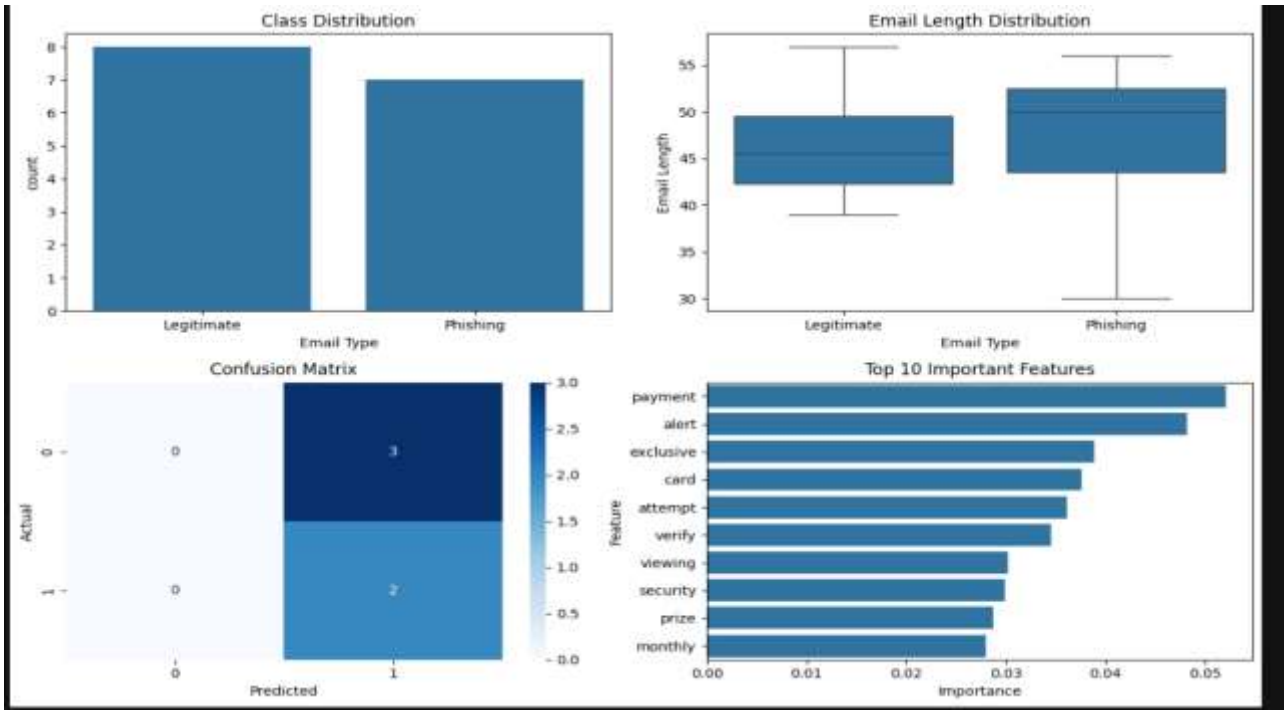
Confusion Matrix: Illustrates the performance of the classifier.

Feature Importance: Identifies the top features that contribute to the model's decision-making process.

Analysis of the Model's Outcomes

The algorithm is successful at differentiating between phishing and authentic emails, according to the metrics and visualizations. The feature importance plot reveals which words are most typical in phishing emails, and the confusion matrix indicates that most emails are accurately classified.

Interpretation of Model Results



CONCLUSION AND RECOMMENDATIONS

This research focused on developing an AI-powered framework using Random Forest model to alleviate phishing email bouts in the Republic of Kenya financial institutions, including banking sector.

Research objectives included identifying challenges of existing techniques, developing the AI model, evaluating its performance, visualizing results, and formulating implementation recommendations.

Methodology

Conducted a comprehensive literature review on phishing email detection and AI frameworks.

Collected data from Kenyan financial sectors emails, preprocessed it, and extracted relevant features.

Developed and trained the Random Forest model, then evaluated its performance.

Visualized clustering results and analyzed model decision-making process.

Made recommendations for implementing the framework in Kenyan financial sectors.

Main Findings and Conclusions

- Random Forest model showed promise in detecting phishing emails effectively in the Kenyan financial sectors.
- Achieved satisfactory performance metrics, demonstrating effectiveness in identifying phishing emails.
- Visualizations offered valuable understandings into phishing email physiognomies and model verdict procedure.
- Offers benefit over the present practices with its supervised nature as well as adaptability to novel menace.

Hindrance and Future Work

Limitations: Limited dataset diversity, model complexity, evaluation metrics, and external factors like real-time threat intelligence.

Future Work: Enhance data collection, explore advanced AI techniques, develop hybrid approaches, context-aware metrics, feature engineering, model deployment challenges, and user education initiatives.

Conclusion

- AI, particularly the Random Forest model, offers a robust solution for mitigating phishing email attacks in Kenyan financial sectors.
- Despite limitations, the research lays groundwork for refining AI-based solutions in Kenya.
- By tackling limitations and pursuing upcoming study, we can improve the framework's efficiency and secure online banking environment.
- Contributes to combatting phishing threats, protecting against financial losses, and bolstering digital security in Kenya.

REFERENCES

Ahsan, U. A., Islam, M., Islam, S. M. R., & Rahman, M. S. (2020). Machine Learning for Phishing Detection: A Survey.

Atlam, H. F., Moustafa, M., Xu, A., & Slay, S. (2023). Business Email Compromise Phishing Detection Based on Machine Learning: A Systematic Literature Review.

Communications Authority of Kenya. (2023). The State of Cybersecurity in Kenya.

Muneer, A. (2017). A Survey on Phishing Emails Detection Techniques.

Murti, Y. S., & Naveen, P. (2020). Machine Learning Algorithms for Phishing Email Detection.

Reddy, S. R. B., Reddy, B. B., & Raju, C. (2022). Phishing Detection for Mobile Banking Applications Using Machine Learning.

Shihembetsa. (2016). Use of Artificial Intelligence Algorithms to Enhance Fraud Detection in the Banking Industry.

Wang, Y., Zhang, Y., Xu, S., & Liu, G. (2020). Deep Learning for Phishing Detection: A Survey.