Defending Mobile Financial Users: Smishing Attack Prevention and Classification Through Machine Learning

N. R. THIRUMALAZGHAGAN[1], E. R. RAMESH[2]

[1] *M.Sc. Department of Computer Science and Engineering, Dr. MGR Educational and Research Institute, Chennai, India*

[2] *Faculty, Centre of Excellence in Digital Forensics, Dr. MGR Educational and Research Institute, Chennai, India*

Abstract— Half of all daily transactions take place on the African continent, and by the end of 2021, it is expected that the value of all transactions worldwide will have reached $3 billion. Phishing incurs losses in the millions for both people and organizations. Smishing is a tactic used to fool mobile money system owners into sending virtual currency to their phones. The context and subtleties involved in gathering the attack's specifics are the only distinctions between phishing and smishing. This research introduces a machine learning-driven approach for categorizing smishing messages aimed at mobile money users. The findings demonstrate that a combination of Extratree classifier feature selection and Random Forest, employing TFIDF vectorization, produces the most effective model, achieving an accuracy score of 99.86%. Smishing gets harder to spot as a consequence. By leveraging minimal resources vocabulary, several models and concepts for detecting smishing crimes were constructed. A machine-learning paradigm can be used to classify texts that have been smuggled.

Index Terms- Natural language processing, Machine-learning algorithm (ExtraTree Classifier), Mobile money, SMS, Social engineering, Smishing.

I. INTRODUCTION

Swahili, a language stemming from the Swahili community, is known as the dominated language below the Sahara. It has official recognition in key African groups like the African Union (AU), the Southern African Development Community (SADC), and the East African Community (EAC). Used in over 16 African nations, Swahili is also a common language among the Indian Ocean coast from Somalia to Mozambique. It extends into parts of Zambia, Malawi, South Africa, Comoros, Botswana, and The Democratic Republic of Congo. Roughly 30–40% of its vocabulary is borrowed from non-Bantu languages, mainly Arabic and Persian, making Swahili a popular Bantu dialect [1]. It ranks in the top 10 most spoken languages globally, which has 200 million speakers. Despite Swahili's prevalence, many of the world's 7000+ languages lack sufficient digital processing data, leading to a distinction between "low-resourced" and "high-resourced" languages [2]. These low-resourced languages, such as Swahili, Bengali, and Punjabi, are used by millions but face limited research, few data sources, and scarce computational tools.

In the context of financial inclusion, particularly in lower-middle-income countries, a significant portion of the population doesn't have access to formal financial services, necessitating more inclusive models. Kenya introduced a solution known as "MPesa," which utilizes mobile numbers as wallets to offer financial services. This mobile money ecosystem, involving customers, agents, and mobile network operators, supports various transactions and has seen substantial growth, particularly in Sub-Saharan Africa. [2]

However, the rise in mobile money platforms has attracted cyber attackers. Smishing is also a type of phishing attack using SMS, influencing social engineering to trick users into sharing confidential information due to the trust associated with SMS communication and its high success rate compared to email spam.

These attacks have known to be challenging. Traditional techniques like rule-based systems and blacklisting are less effective in developing attack strategies. Machine-learning algorithms are more effective in identifying and addressing these threats, as illustrated by progress in message classification.

This research is inspired by these advanced and proposed machine-learning models to categorize Swahili Smishing messages aimed at mobile money users. Using real-world data from Tanzania, the study aims to introduce an efficient detection system to protect users from financial losses caused by social engineering attacks.

The remaining sections of the paper detail related research, objectives, methodologies, results, and discussions, concluding with suggestions for future studies and interventions in fighting cyber threats targeting vulnerable populations.

II. REVIEW OF LITERATURE

Saleem Raja Abdul Samad, Pradeepa Ganesan, et al., [3] had proposed SMS is crucial for quick text communication, yet it inadvertently enables smishing due to its ubiquity and reliability. Smishing attackers exploit this trust to deceive users into sharing sensitive information. Early detection is vital. Researchers propose using Machine Learning and Language Processing to distinguish legitimate from fraudulent SMS. This paper introduces two methods, SmishGuard, leveraging Ml models and language processing. Results show TF-IDF with LDA outperforms weighted Average Word2Vec, while Random Forest and Extreme Gradient Boosting achieve higher accuracy in detecting smishing attacks.

Bodunde Odunola Akinyemi, Dauda Akinwuyi Olalere, et al., [4] had proposed various classifier to predict cyber threats in mobile money service onboarding. Six supervised ML algorithms are tested, including Logistic Regression, Navies Bayes, SNN, DNN, CART, and RF, with different setups. Models are evaluated with and without SMOTE on a dataset of 25,000 mobile money applications. Random Forest with SMOTE performs best, especially in multiclass scenarios, Enhancing security by detecting fraudulent registration during mobile money onboarding for the unbanked.

Resenthiran Kohilan, Harsha Edirisinghe Warakagoda, et al., [5] had proposed Smishing attacks via fraudulent text messages are widespread, targeting personal data like password and credit card details, leading to identity theft and financial loss. This research emphasizes the need for strong prevention measures. It explores new prevention methods, evaluates current protection systems, and investigates adding smishing detection to social media platforms. Various ML models are trained and tested on gathered datasets, with CNN achieving the highest accuracy. The study contributes significantly to smishing prevention with user-friendly features and multilingual detection, ensuring a secure online environment.

Rubaiath E. Ulfath, Iqbal H. Sarker, et al., [6] had proposed Phishing scams via SMS are rampant due to smartphone prevalence and mobile internet. AI-powered cyber security faces challenges in detecting phishing SMS. This paper explores ML and NLP to distinguish phishing from legitimate texts. SVM, after features extraction and selections, achieves 98.39% accuracy, with 98.27% cross-validation and 99.0% F1score for legit SMS. Tested methods are evaluated using standard metrics on a benchmark dataset.

David Njuguna, John Kamau and Dennis Kaburu, [7] had proposed Cybercrimes is on the rise due to technological advancements and increased smartphone reliance. In Kenya, smishing (SMS Phishing) attacks have surged, but no comprehensive investigation has been conducted. Existing solutions lack sender authentication, content filtering, and user notification capabilities. This study introduces a novel method using Python, MySQL, and Navies Bayes classifier to address these shortcomings, aiding users in quickly identifying and mitigating fraudulent smishing messages.

Fillemon S. Enkono and Nalina Suresh, [8] had proposed fraudulent e-wallet deposit notification SMS are widespread in Namibia, targeting mobile banking users to steal money and goods. To address this, a study assessed machine learning’s effectiveness in detecting these scams. Navies Bayes (NB) and support vector machine (SVM) classifiers were trained to distinguish between legitimate and fraudulent SMS. The evaluation showed that the SVM classifier outperformed NB in efficiently identifying fraudulent SMSes.

Sandhya Mishra and Devpriya Soni, [9], had proposed ‘smartphones’ popularity and constant internet access have made them vulnerable to phishing and smishing attacks. Phishing involves malicious emails, while smishing combines SMS and phishing, often redirecting users to harmful websites or apps. Existing methods to detect smishing suffer from false positives. To address this, we proposed the 'Smishing Detector', comprising four modules: SMS content Analyzer, URL Filter, Source Code Analyzer, and APK Download Detector. Leveraging Navies Bayes and URL inspection, our prototypes achieve 96.29% accuracy by analyzing text, URLs, source code, and downloaded files, enhancing security compared to existing models.

III. RESEARCH METHODOLOGY

The aim of our research is to identify an effective machine learning algorithm for classifying Smishing messages aimed at mobile financial users. We opted for machine learning models because they require less data and hardware compared to deep learning models [10]. Smishing messages targeting mobile money users typically employ a carefully crafted pattern of words and include a mobile number for receiving electronic money from victims. Fig. 1 presents the overall architecture of the proposed approach. After data collection, messages undergo preprocessing to remove unnecessary words, such as stopwords. Next, tokenization is applied, converting sentences into lists of words, which is essential for the vectorization of text at both the word and character levels [11]. We use word vectorization to reduce the dimensionality of the resulting vectors, employing count and TFIDF vectorizers. During model training, feature selection and parameter tuning are performed. The model is trained using two techniques: bag of words and n-grams, with 2-5 n-grams used to identify the best performing model.

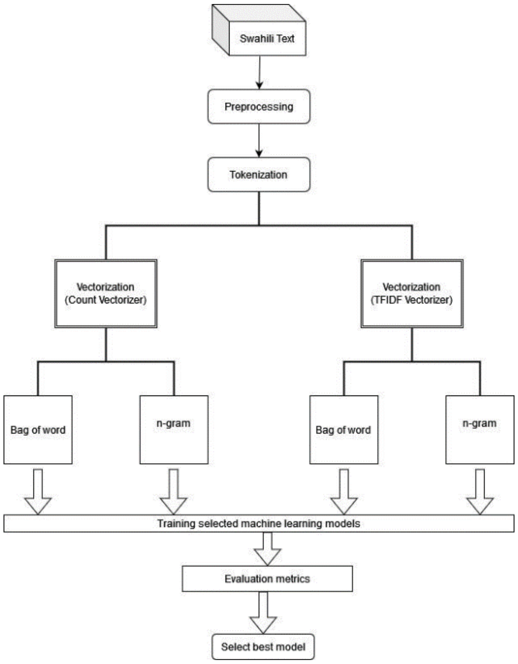


Fig. 1: General architecture of the Smishing filtration model

A. DATA COLLECTION:

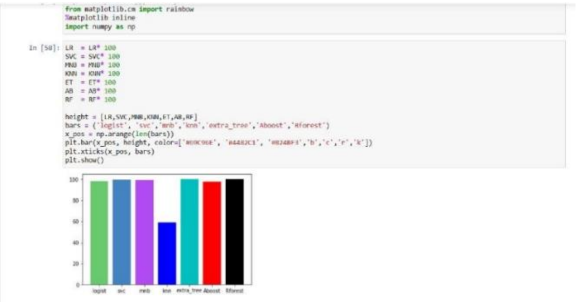
With special permission, we made our dataset available on GitHub. It includes Smishing SMS collected from mobile network operators. On January 25, 2021, we gathered 874,044 Smishing messages, out of which 136 were unique. Another batch collected on January 31, 2021, contained 937,716 messages, with 161 unique ones. Additionally, volunteers contributed five unique Smishing messages [12]. Due to privacy policies, we couldn't obtain legitimate messages directly from mobile operators. Instead, we asked university students to forward messages over five months, from February to June 2021. We adhered to all ethical guidelines, allowing volunteers to choose which messages to share [17]. Ultimately, we collected 31,962 legitimate messages. Since this resulted in an imbalanced dataset, we addressed it by undersampling the majority class, selecting 11,061 legitimate messages for model training.

B. TEXT PREPROCESSING:

Experts manually and consistently labelled the dataset with spam and legitimate tags. We used Python library functions for text preprocessing and data cleaning. This included converting all text to lowercase and removing punctuation marks [13]. Numeric values were retained because they can represent significant amounts to be transferred, evade rule-based detection systems, or indicate a mobile number used by attackers. Stopwords were removed using a predefined list, and the dataset was tokenized to produce a list of words used as input features.

C. ALGORITHM:

After applying data pre-processing techniques, six machine learning algorithms were chosen to implement the model: Multinomial Naïve Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Random Forest, AdaBoost, and ExtraTreeClassifier as shown in *Fig.2*. These models were evaluated using a feature vector created by a count vectorizer [18]. Among these, the Multinomial Naïve Bayes model performed poorly. Classifying smishing (SMS phishing) messages is a particularly sensitive task due to the serious implications of false positives and false negatives [14]



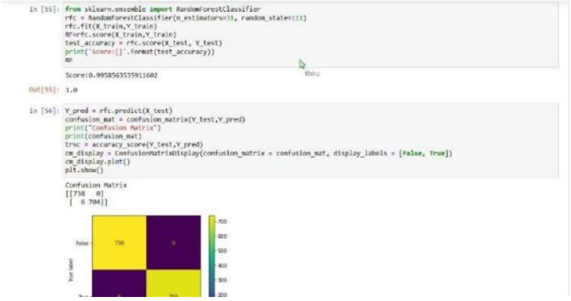
```
from matplotlib.cm import rainbow
%matplotlib inline
import numpy as np
```

```
In [58]: LR  = LR* 100
         SVC = SVC* 100
         MNB = MNB* 100
         KNN = KNN* 100
         ET  = ET* 100
         AB  = AB* 100
         RF  = RF* 100

         height = [LR,SVC,MNB,KNN,ET,AB,RF]
         bars = ('logist', 'svc','mnb','knn','extra_tree','Aboost','Rforest')
         x_pos = np.arange(len(bars))
         plt.bar(x_pos, height, color=['#00C96E', '#4482C1', '#8248F3','b','c','r','k'])
         plt.xticks(x_pos, bars)
         plt.show()
```

Fig.2: List of algorithms used

D. PREDICTION:

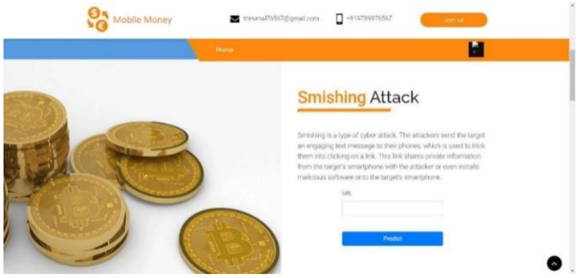
This paper analyzes an appropriate algorithm to classify legitimate messages from Smishing messages targeting mobile money users. We successfully implemented various machine-learning algorithms to find the best fits for this mode [15]. The results from the experiments show that Random Forest evaluates the best accuracy score of 99.86% as shown in *Fig. 3*. Therefore, it can be concluded that a hybrid of the Extratree classifier feature selection technique in conjunction with Random Forest, taking 750 as the maximum number of features vectorized by the TFIDF technique, returns the best accuracy score[16].



```
In [55]: from sklearn.ensemble import RandomForestClassifier
         rfc = RandomForestClassifier(n_estimators=31, random_state=111)
         rfc.fit(X_train,Y_train)
         RF=rfc.score(X_train,Y_train)
         test_accuracy = rfc.score(X_test, Y_test)
         print('Score:{}'.format(test_accuracy))
         RF

         Score:0.9958563535911602

Out[55]: 1.0

In [56]: Y_pred = rfc.predict(X_test)
         confusion_mat = confusion_matrix(Y_test,Y_pred)
         print("Confusion Matrix")
         print(confusion_mat)
         trsc = accuracy_score(Y_test,Y_pred)
         cm_display = ConfusionMatrixDisplay(confusion_matrix = confusion_mat, display_labels = [False, True])
         cm_display.plot()
         plt.show()

         Confusion Matrix
         [[738   0]
          [  6 704]]
```

Fig. 3: Accuracy score of Random forest with hybrid of Extra tree classifier

Fig.4 represents the interface of our website. In the tab below, users can paste messages classified as either "ham" (good messages) or "spam" (bad messages). This classification is determined by the algorithms used in our system.

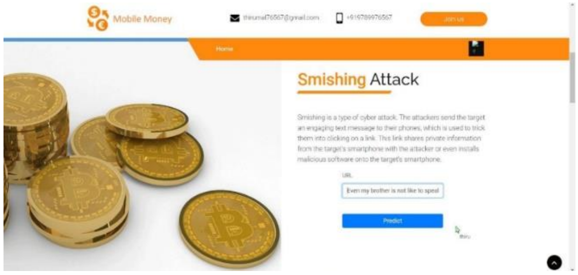


Smishing Attack

Smishing is a type of cyber attack. The attackers send the target an engaging text message to their phones, which is used to trick them into clicking on a link. This link shares private information from the target's smartphone with the attacker or even installs malicious software onto the target's smartphone.

Fig.4: Website interface

Fig.5: represent the result of message entered weather it is ham or spam.



Smishing Attack

Smishing is a type of cyber attack. The attackers send the target an engaging text message to their phones, which is used to trick them into clicking on a link. This link shares private information from the target's smartphone with the attacker or even installs malicious software onto the target's smartphone.

URL: Even my brother is not like to speal

Fig.5: Result of message entered

CONCLUSION

Recently, mobile network operators have been experiencing a significant surge in Smishing attacks. These attacks come in two forms: general attacks affecting a wide range of users and targeted attacks specifically aimed at mobile money users. Given the push for financial inclusion through mobile money in the East African region, Smishing attacks targeting these users have become alarmingly common. In response to this growing threat, our paper focused on identifying an effective algorithm for distinguishing between legitimate messages and Smishing messages targeting mobile money users. We conducted thorough investigations into various machine learning algorithms to determine which one best suited our specific context. The results of our experiments revealed that Random Forest achieved the highest accuracy score, reaching an impressive 99.86%. This indicates its effectiveness in accurately classifying messages. Moreover, we found that combining the Extratree classifier feature selection technique with Random Forest, and setting the maximum number of features vectorized by the TFIDF technique to 750, yielded the best accuracy score. Looking ahead, we plan to leverage these findings to develop a mobile application that utilizes the identified algorithm to protect users from Smishing attacks. Additionally, we aim to explore deep learning methodologies to further enhance our approach. By doing so, we hope to minimize the occurrence of false positives and false negatives, which can have significant consequences for users, ranging from financial losses to overlooking important messages.

REFERENCES

[1] Y. Lodhi, Oriental Influences in Swahili. A Study in Language and Cultural Contacts. 2000.

[2] J. Hirschberg and C. D. Manning, “Advances in natural language processing,” Science, vol. 349, no.6245, pp. 261–266, 2015.

[3] Saleem Raja Abdul Samad, Pradeepa Ganesan, Justin Rajasekaran, Madhubala Radhakrishnan, Hariraman Ammaippan and Vinodhini Ramamurthy, “SmishGuard: Leveraging Machine Learning and Natural Language Processing for Smishing Detection” International Journal of Advanced Computer Science and Applications(IJACSA), 14(11), 2023.

[4] Akinyemi, Bodunde Odunola; Olalere, Dauda Akinwuyi; Sanni, Mistura Laide; Olajubu, Emmanuel Ajayi; Aderounmu, Ganiyu Adesola; et al.”Performance Evaluation of Machine Learning Models for Cyber Threat Detection and Prevention in Mobile Money Services” Informatica; Ljubljana Vol. 47, Iss. 6, (May 2023): 173-190. DOI:10.31449/inf.v47i6.4691.

[5] R. Kohilan, H. E. Warakagoda, T. T. Kitulgoda, N. Skandhakumar and N. Kuruwitaarachchi, "A Machine Learning-based Approach for Detecting Smishing Attacks at End-user Level," 2023 IEEE International Conference on e-Business Engineering (ICEBE), Sydney, Australia, 2023, pp. 149-154, doi: 10.1109/ICEBE59045.2023.00042.

[6] Ulfath, R.E., Sarker, I.H., Chowdhury, M.J.M., Hammoudeh, M. (2022). Detecting Smishing Attacks Using Feature Extraction and Classification Techniques. In: Arefin, M.S., Kaiser, M.S., Bandyopadhyay, A., Ahad, M.A.R., Ray, K. (eds) Proceedings of the International Conference on Big Data, IoT, and Machine Learning. Lecture Notes on Data Engineering and Communications Technologies, vol 95. Springer, Singapore.

[7] D. Njuguna, J. Kamau and D. Kaburu, "Model For Mitigating Smishing Attacks On Mobile Platforms," 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa, 2021, pp. 1-6, doi: 10.1109/ICECET52533.2021.9

[8] ENKONO, Fillemon S. and SURESH, Nalina. Application of Machine Learning Classification to Detect Fraudulent E-wallet Deposit Notification SMSes. AJIC [online]. 2020, vol.25 [cited 2024-05-17], pp.1-12.

[9] Sandhya Mishra, Devpriya Soni, Smishing Detector: A security model to detect smishing through SMS content analysis and URL behavior analysis, Future Generation Computer Systems, Volume 108, 2020, Pages 803-815, ISSN 0167-739X,

[10] A. Y. Lodhi, Oriental Influences in Swahili. A Study in Language and Cultural Contacts. 2000.

[11] B. E. Coleman, “A history of Swahili,” Black Sch., vol. 2, no. 6, pp. 13–25, 1971.

[12] UNESCO, “World Kiswahili Language Day,” in 41st Session, Paris, 2021, vol.41 C/61. Accessed: Jan. 29, 2022. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000379702

[13] S. M. Lakew, M. Negri, and M. Turchi, “Low

resource neural machine translation: A benchmark for five African languages," ArXiv Prepr. ArXiv200314402, 2020.

[14] A. Magueresse, V. Carles, and E. Heetderks, "Low-resource Languages: A Review of Past Work and Future Challenges," ArXiv200607264 Cs, Jun. 2020, Accessed: Jan. 29, 2022. [Online]. Available: http://arxiv.org/abs/2006.07264

[15] J. Hirschberg and C. D. Manning, "Advances in natural language processing," Science, vol. 349, no. 6245, pp. 261–266, 2015.

[16] C. Cieri, M. Maxwell, S. Strassel, and J. Tracey, "Selection criteria for low resource language programs," in Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016, pp. 4543–4549.

[17] A. K. Singh, "Natural Language Processing for Less Privileged Languages: Where do we come from? Where are we going?" 2008.

[18] Y. Tsvetkov, "Opportunities and challenges in working with low-resource languages," Slides Part-1, 2017.