International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 434-440 www.ijaem.net ISSN: 2395-5252

Machine Learning Based Online Scam Reportage System for Security Agencies

O. K. Akinyokun[1], T. F. Owolabi[2]
[1,2]*Department of Cyber Security, Federal University of Technology,Akure, Nigeria.*

Date of Submission: 20-10-2024 Date of Acceptance: 30-10-2024

ABSTRACT: The use of technology in our daily lives has led to an increase in cybercrime worldwide. Recently, cybercrime cases have been on the rise. This project aimed to develop a mobile application for reporting online scams aiding machine learning. The mobile application allows users to simultaneously report cybercrime incidents to security agency, attach evidence to the report.
The application was developed using flutter, dart, scikit-learn, python and the development process followed the agile development methodology. The application was tested for functionality, usability, and security, and it was found to be fully functional, highly usable, and secure. Machine learning was integrated with the mobile app to analyze collected reports and identify patterns in scammer tactics, communication styles, and potential red flags. Random forest classifier was used to train the model. The Random Forest Classifier performed well in detecting scam communications, achieving an accuracy of 91.12% during testing, key metrics used to evaluate the model include precision, accuracy, recall, and F1 score, all of which demonstrated the system's effectiveness in identifying online scams. The mobile application for reporting online scams has the potential to improve the reporting and investigation of cybercrime in Nigeria.
KEYWORDS: Machine Learning, Random Forest Classifier, Online Scam, Scam Detection, Reportage System, Cyber security.

I. INTRODUCTION

Online scams are becoming a global concern, and security agencies are increasingly challenged in tackling them. A scam is a fraudulent scheme or deception aimed at extorting money or obtaining personal information from unsuspecting victims. With the prevalence of internet access and the rapid advancement of technology, scammers are utilizing various methods to defraud their victims online. To combat this issue, there is a need to develop an android-based online scam reportage system using machine learning to help users and security agencies detect and prevent these scams.

The majority of the time, the process of reporting and investigating online scams in society has been manual. Most Individuals with complaints go to their various banks of operation (in financial cases) or any security agent's office (e.g. Nigerian Police stations) to notify and write down matters that require their attention. These officers will raise an incident form in these offices and ask the reporter to fill out the required sections.

As a result, the mobile strategy is the most cost-effective and open way to report online scam, with far-reaching advantages and coverage. Despite various global efforts in this field, Nigeria still lacks a widely adopted and effective system for reporting online scams.

Online scams have been steadily increasing over the years, with millions of people becoming victims each year. According to Palad et al. (2019), using data mining techniques to analyze reports of online scam incidents can help uncover patterns and trends, which can be really useful in both spotting and stopping these scams. Similarly, Bilz et al. (2023) also carried out a systematic literature review on online romance scams, to provide insight on the nature of these scams and the methods used to carry them out. Their research emphasizes the need for a reliable system that can effectively gather and analyze data to help prevent people from becoming victims of online scams.

The emotional harm and financial loss of online scams are colossal, as victims not only lose their hard-earned money but also endure serious psychological distress. Aborisade et al. (2023) explored the experiences of victims of romance scams in Nigeria and discovered that many of them struggle with psychological issues like depression, anxiety, and a deep sense of mistrust. Additionally, Ravenelle et al. (2022) emphasized the importance of detecting, normalizing, and addressing online

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 434-440 www.ijaem.net ISSN: 2395-5252

scams that surged during the COVID-19 pandemic. Their research highlights the serious need for a robust reporting system to help prevent individuals from falling victim to these scams.

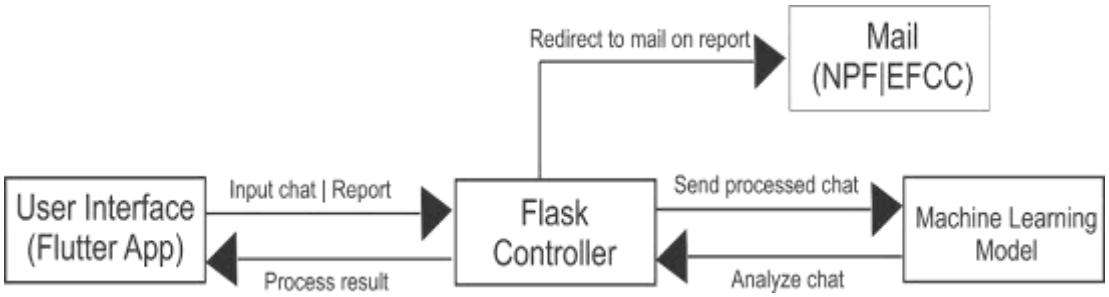
Verma et al. (2020) conducted an analysis of patterns found on online scam websites, and pointed out certain features that are common across these sites. Their findings shows the importance for stakeholders to recognize these components to help prevent people from becoming victims of online scams. Lee et al. (2022) also took carried out their research to look at profiles of scammers and regular users on online dating platforms. They identified specific linguistic patterns and descriptions that scammers tend to use. These studies can be invaluable in developing effective strategies to detect and prevent online scams.

With technology advancing rapidly, we have more opportunities to create efficient systems for reporting online scams. For instance, artificial intelligence and machine learning algorithms can play a key role in spotting and analyzing online scams. Calderon et al. (2020) utilized decision tree algorithms to classify data related to online scams in the Philippines. Their algorithm demonstrated high accuracy in identifying different types of scams, which can help in developing a systematic approach to combat these issues. Machine learning has made significant strides in various fields, such as healthcare, finance, and cybersecurity, enhancing decision-making processes. In this research, we'll be systems focused on cybersecurity. This system will play a crucial role in the fight against online scams, helping to reduce the emotional and financial burdens faced by victims. Overall, the significance of this research is substantial, as it contributes to creating technology-driven solutions to address pressing societal challenges.

User Interface (Flutter App): This component serves as the front-end of the application, developed using Flutter for a cross-platform experience. Users interact with this interface to submit chat data suspected of containing scam activities. After processing and analysis, the results are displayed here, allowing users to view the analysis and decide on further actions.

Flask Controller: Acting as the middleware, the Flask Controller receives the input chat or report from the User Interface. It handles the requests byprocessing the data and coordinating between the User Interface and the Machine Learning Model. Additionally, if a report needs to be sent, it facilitates the redirection to the user's mail system for report submission.

II. SYSTEM ARCHITECTURE



using machine learning to analyze data gathered from different sources to identify potential online scams. The system will then notify security agencies, giving them detailed reports about the scams detected. By incorporating machine learning, we can boost the system's accuracy and its ability to detect and prevent online scams effectively.

The proposed Android-based scam reporting system utilizing machine learning aims to equip security agencies with a powerful tool to tackle the increasing number of online scams. This system will analyze incoming data to flag potential scams and alert security agencies, complete withcomprehensive reports on these incidents.

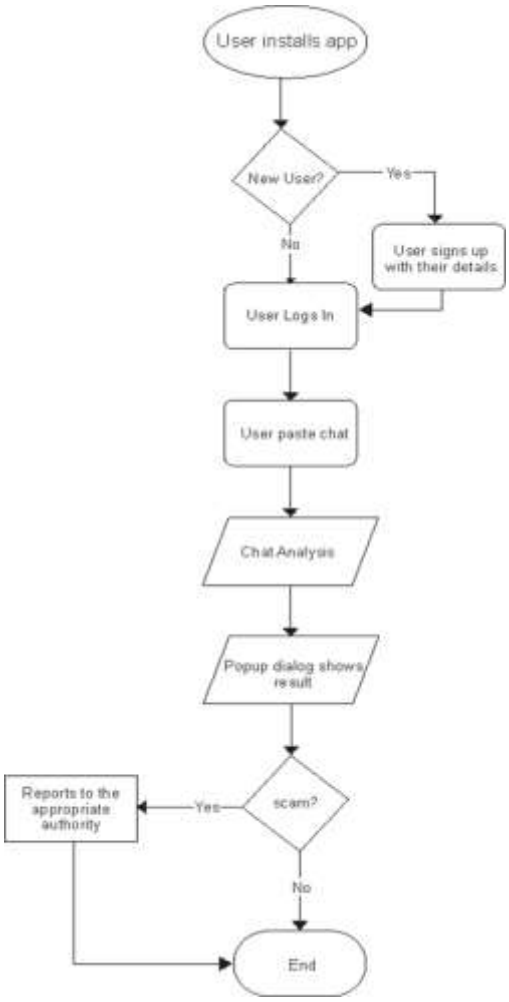
The insights gained from this research willgive way for developing future Android-based

Machine Learning Model: This backend component is where the analysis of the submitted chat data takes place. Using machine learning algorithms, it assesses the content to determine potential scam activities. Based on its analysis, it sends back the processed data to the Flask Controller, which includes a determination of whether the interaction is a red flag.

Mail: This component handles external actions like opening the user's default mail client. When the Flask Controller determines a report needs to be made based on the analysis, it triggers the Mail component, which redirects users to their email

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 434-440 www.ijaem.net ISSN: 2395-5252

client, pre-populating the mail with the necessary report details to be sent to the security agency.



ACTIVITY DIAGRAM OF THE ONLINE SCAM REPORTAGE SYSTEM

The mobile application enables users to report crimes in panic mode way. Here user will be able to login to Mobile App and give brief description or paste chats and press a button on the app to analyze the pasted chats for scams and then report the case if true, and then the security agencies will be able to collect the report and analyze it for further investigation.

III. REQUIREMENTS FOR DEVELOPMENT

Establishment of the system requirements is very important in the process of Mobile app development, it guides the design and functionality of the application. The requirements for developing the proposed application are outlined.

Choice of programming language

The following tools are necessary for the development of the proposed solution:

a. python
b.dart

Hardware Requirements

These are physical components and resources necessary for developing the online scam reportage system:

Any computer with at least;
a.50 GB Hard disk space
b.2GB Ram

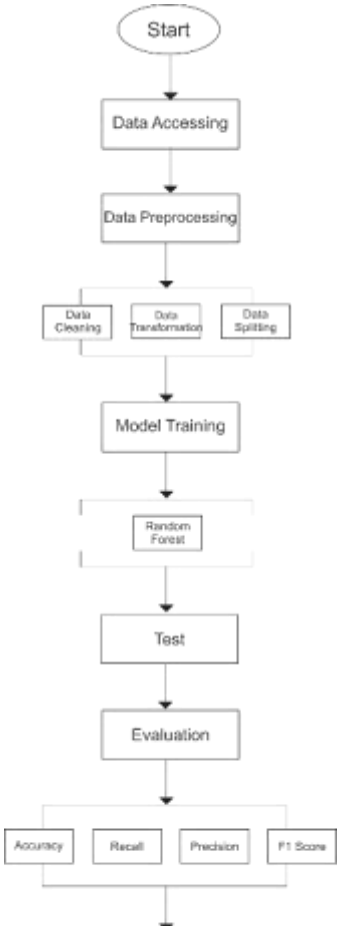
Software Requirements

The following software platforms are necessary for the development and deployment of the proposed Application:

a. Windows 7 and above
b.Independent Development Environment (IDE): A development environment is necessary to develop

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 434-440 www.ijaem.net ISSN: 2395-5252

the Application. In this project, Android Studio was used. Necessary steps to setup Android studio development environment includes downloading and installing Android Studio from the official website: https://developer.android.com/studio.



MODEL TRAINING PROCESS OF THE SCAM DETECTOR SYSTEM

DATA ACCESSING

The accessing phase of the proposed model involves collecting data through various methods, such as surveys, interviews, observations, and experiments. The specific methods used will depend on the type of research being conducted and the research questions being asked. The process involves loading the dataset, checking the shape, duplicates, missing values.

The training data is imported using a python library called pandas. Pandas is an open source library used by data enthusiasts and machine learning engineers to handle complex data structures like spreadsheets and CSV files.

DATA PREPROCESSING

The training data was preprocessed using the Python, in the following steps:

1. Data Cleaning

a.Handle Missing Values: The pandas library was used to check for rows with missing data, then those rows were deleted from the training data to reduce noise and increase efficiency in the model.

b. Remove Duplicates: The pandas library was also used to identify and delete duplicate records to avoid redundancy in the dataset.

c.Using the re library Python was used to write regular expression filters to convert all texts to lower case as well as remove stopwords and unnecessary symbols in the training data.

2. Data Transformation

a.Label Encoding: Categorical values in the dataset were converted into numerical labels of 1s and 0s for spam and non-spam texts respectively.

3. Data Splitting

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 434-440 www.ijaem.net ISSN: 2395-5252

a.Train-Test Split: The Python library scikit-learn was used to divide the data into training and testing sets, typically following a 90-10 ratio.

ALGORITHM TRAINING

This section provides a thorough step-by-step process of how the algorithm was developed to detect fraudulent texts. Random forest was desired for the training to avoid overfitting and it achieved a 91.12% accuracy.

RANDOM FOREST CLASSIFIER

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is controlled with the max samples parameter if bootstrap=True (default), otherwise the whole dataset is used to build each tree.

For each decision tree in a random forest algorithm, each node's importance is calculated. The formula below assumes there are two features being used.

$$ni_j = w_j C_j - W_{left\ (j)} C_{left\ (j)} - W_{right\ (j)} C_{right\ (j)} \quad (1)$$

niis the importance of node j, w_j is the weighted number of samples reaching node j, $C_{j\ is}$ the impurity value of node j, left(j) is the child node from left split on node j, and right(j) is the child node from right split on node j.

The relevance of each column in the decision tree is then calculated as:

$$fi_i = \frac{\sum_{j:nodejsplitsonfeaturei} ni_j}{\sum_{k \in allnodes} ni_k} \quad (2)$$

fi_i is the importance of feature and $ni_j$is the importance of node j.

DATA SOURCE

The data for this research was gotten from kagglehttps://www.kaggle.com/datasets/team-ai/spam-text-message-classification.

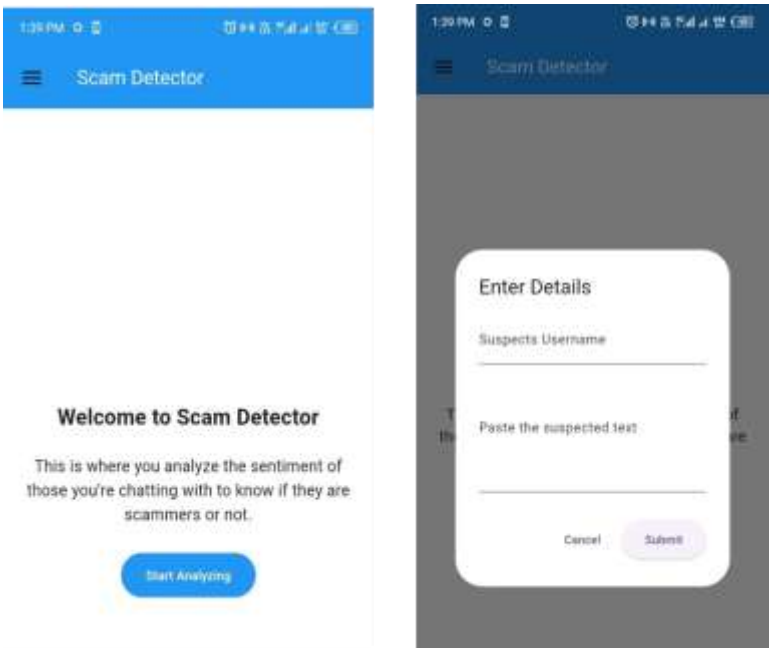
Kaggle is a platform used for data science and machine learning. It brings about a vast repository of open-source data covering various domains like healthcare, finance, social media, and more. Kaggle provides interactive Jupyter notebooks, allowing the writing and sharing code for data analysis and model building. This research used Jupyter notebook for the coding of this algorithms.

IV. IMPLEMENTATION OF APPLICATION

The Implementation phase is a very critical phase during development of any Mobile App development. This Application was designed with usability and efficiency in mind, ensuring that users can navigate the application easily to analyze chats and report online scams to security agencies easily.

The Application users needs to register and login, with appropriate authentication measures, in this case email and password. Users are able to report online scams to security agencies. During the online scam reporting process, users are provided with form fields to guide the user during through the reporting process. When all required fields are filled by the user, the application allows users to report, and then reporting application takes the user to the email application for final submission to the security agencies.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 434-440 www.ijaem.net ISSN: 2395-5252



HOME PAGE OF THE ONLINE SCAM REPORTAGE SYSTEM

```
In [25]:
from sklearn.ensemble import RandomForestClassifier

rfc=RandomForestClassifier()

rfc.fit(X_train, y_train)

rfc.score(X_test,y_test)

Out[25]:
0.9112107623318386

In [26]:
from sklearn.metrics import precision_score, recall_score, f1_score, accuracy_score

y_pred = rfc.predict(X_test)

# Calculating Precision, Recall, F1 Score, and Accuracy
precision = precision_score(y_test, y_pred, average='micro')
recall = recall_score(y_test, y_pred, average='micro')
f1 = f1_score(y_test, y_pred, average='micro')
accuracy = accuracy_score(y_test, y_pred)

# Formatting the results to 2 decimal places
formatted_precision = "{:.2f}".format(precision*100)
formatted_recall = "{:.2f}".format(recall*100)
formatted_f1 = "{:.2f}".format(f1*100)
formatted_accuracy = "{:.2f}".format(accuracy*100)

print(f'Precision: {formatted_precision}')
print(f'Recall: {formatted_recall}')
print(f'F1 Score: {formatted_f1}')
print(f'Accuracy: {formatted_accuracy}')

Precision: 91.12
Recall: 91.12
F1 Score: 91.12
Accuracy: 91.12
```

CODE SNIPPET AND RESULT FOR MODEL EVALUATION

V. CONCLUSION

In this project, a mobile application that can be used to detect online scam and report those cases and address the need of users with difficulty in reporting online scams, was developed. Thus, the mobile application comes appropriate to provide a solution not only for users to detect online scam messages but to also create way for users to report these fraudulent cases.

In conclusion, this project developed a mobile application for simultaneously for detecting online scam messages and reporting fraudulent cases to security agencies in Nigeria. The application was developed using Dart programming language,

Python, Flutter, and the development process followed the Agile Methodology. The application was tested for functionality, usability, and security, and it was found to be fully functional, highly usable, and secure.

REFERENCES

[1]. Abdelhamid, N., Ayesh, A., &Thabtah, F. (2014). Phishing detection: A case analysis of classifiers' performance. International Journal of Information Security, 13(2), 142–153. https://doi.org/10.1007/s10207-013-0204-0

[2]. Abdelhamid, N., Ayesh, A., &Thabtah, F. (2014). Phishing detection based associative classification data mining. Expert Systems with Applications, 41(13), 5948-5959. https://doi.org/10.1016/j.eswa.2014.03.019

[3]. Alameri, T., Alhilali, A. H., Ali, N. S., &Mezaal, J. K. (2022). Crime reporting and police controlling: Mobile and web-based approach for information-sharing in Iraq. Journal of Intelligent Systems, 31(1), 726–738. https://doi.org/10.1515/jisys-2022-0034:contentReference[oaicite:0]{index=0}:contentReference[oaicite:1]{index=1}.

[4]. Alharbi, H., Alharbi, A., &Alharbi, H. (2019). Investigating the Factors Influencing the Adoption of Cybercrime Reporting Mobile Applications. Journal of Cybersecurity, 5(1), e10.

[5]. Alnemari, S., &Alshammari, M. (2023). Detecting phishing domains using machine learning. Applied Sciences, 13(8), 4649. https://doi.org/10.3390/app13084649

[6]. Basnet, R., Mukkamala, S., & Sung, A. H. (2008). Detection of phishing attacks: A machine learning approach. Studies in Fuzziness and Soft Computing, 226, 373–383. https://doi.org/10.1007/978-3-540-71096-6_23

[7]. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

[8]. Dabhere, A., Kulkarni, A., Kumbharkar, K., Chhajed, V., &Tirth, S. (2015). Crime area detection and criminal data record. International Journal of Computer Science and Information Technologies, 6(1), 510-513. Retrieved from http://www.ijcsit.com

[9]. Kn, J. &Perera, P. (2021). Impact of Crime Reporting System to Enhance Effectiveness of Police Service. International Journal of Computer Trends and Technology. 69. 1-5.

[10]. Lee, K. F., Chan, M. Y., & Ali, A. M. (2023). Self and desired partner descriptions in the online romance scam: A linguistic analysis of scammer and general user profiles on online dating portals. Crime Prevention and Community Safety, 25(1), 20-46. https://doi.org/10.1080/14658811.2022.2164460

[11]. Mohamed, A., &Abdallah, M. (2021). Machine learning techniques for phishing detection: Review and performance evaluation. Journal of Information Security and Applications, 59, 102805. https://doi.org/10.1016/j.jisa.2021.102805

[12]. Ravenelle, A. J., Janko, E., & Kowalski, K. C. (2022). Good jobs, scam jobs: Detecting, normalizing, and internalizing online job scams during the COVID-19 pandemic. New Media & Society, 24(7). https://doi.org/10.1177/14614448221099223

[13]. Toolan, F., &Carthy, J. (2010). Feature selection for spam and phishing detection. eCrime Researchers Summit (eCRS). https://doi.org/10.1109/ecrime.2010.5706671

[14]. Verma, S., Prabha, D., & Rajput, S. (2020). Detecting phishing websites using machine learning techniques. PLOS ONE, 15(12), e0258361. https://doi.org/10.1371/journal.pone.0258361

[15]. Weijer, S., Leukfeldt, E., &Bernasco, W. (2018). Determinants of reporting cybercrime: A comparison between identity theft, consumer fraud, and hacking. European Journal of Criminology. 16. 147737081877361. 10.1177/1477370818773610.