Scam Call Detection Using NLP and Naïve Bayes Classifier

[1].C. Valarmathi, Assistant Professor, Department of Computer Science and Engineering, Sri Sairam College of Engineering, Anekal, Bangalore,vinmathi20@gmail.com.

[2].S. Sharanya, Student, Department of Computer Science and Engineering, Sri Sairam College of Engineering, Anekal, Bangalore, sridharan784@gmail.com

ABSTRACT

Financial fraud, particularly credit card fraud, is a pressing concern in the realm of digital transactions. The number of phone scams is increasing daily as con artists use phone calls to target victims for nefarious ends. Individuals are falling for con artists' proposals, becoming victims and giving up their personal information, leaving them open to abuse. Effective detection techniques are becoming more and more necessary. In this study, we offer an efficient approach to scam call identification utilizing speech-to-text libraries and the machine learning technique Naïve Bayes classifier. Our technology, which translates voice to text, uses this text to evaluate conversations in real time. It looks for trends and suspicious phrases that point to attempted scams, including asking for credit card numbers, passwords, or other sensitive information. The user will be able to decide whether or not to trust and continue with the call by using the alert prompt that appears as a pop-up message if the words are found to be suspicious. The user will take certain measures, such as ending the conversation right away, blocking the number, and reporting it further, if they don't trust the call. Our strategy is to successfully handle scam calls through ongoing adaptation and learning, boosting user security and confidence in phone conversations. The user will be able to decide whether or not to trust and continue with the call by using the alert prompt that appears as a pop-up message if the words are found to be suspicious. The user will take certain measures, such as ending the conversation right away, blocking the number, and reporting it further, if they don't trust the call. Our strategy is to successfully handle scam calls through ongoing adaptation and learning, boosting user security and confidence in phone conversations.

Keyword: Spam Detection, Naïve Bayes, Natural Language Processing, Machine Learning.

1. INTRODUCTION

Scam calls have become a significant issue in today's society, causing harm to individuals both financially and emotionally. These scams, utilizing tactics such as impersonation, spoofing, and digital manipulation, aim to obtain sensitive information, extort money, or cause harm to person's reputation. These fraudulent activities not only cause financial losses but also erode trust in communication networks. Conventional approaches for detecting fraudulent phone calls typically require humans to spend a lot of time manually reviewing call logs and recordings to detect any suspicious patterns. However, this process is slow, costly, and might not always detect new types of scams accurately. In response to this challenge, there is a growing demand for advanced technological solutions capable of dynamically analysing scam calls in real-time to mitigate their impact.

To address this problem, the development of a scam call tone analyser with real-time data using Natural Language Processing (NLP) and machine learning algorithm (Naive Bayes) has gained attention. This innovative solution aims to analyse the tone of incoming calls in real-time by converting speech to text and identify if the caller is attempting a scam. With the user's permission, the system can terminate the call upon detecting specific scam-related sentences like asking for OTP, CVV, passwords, etc.

The utilization of NLP and machine learning algorithm (Naive Bayes) is crucial in achieving an accurate and efficient scam call tone analyser. NLP techniques allow for the interpretation and understanding of human language by the system. By leveraging NLP, the analyzer can identify patterns and indicators of scam attempts based on the

caller's tone, speech patterns, and choice of words.

Furthermore, the integration of machine learning algorithm (Naive Bayes) enhances the system's ability to classify incoming calls as either legitimate or scam. The algorithm is trained on a large dataset of 6000 scam call and genuine call examples, enabling it to learn and recognize common scam patterns. Through this learning process, the system becomes more accurate at identifying potential scam calls and reducing false positives.

The real-time aspect of the scam call tone analyzer is crucial for providing immediate protection to users. As soon as an incoming call is detected, the system starts analyzing the caller's tone and using NLP techniques to assess the content of the conversation. If specific scam- related sentences are identified, and the user has given permission, the call is promptly terminated. This real- time response minimizes the risk of individuals falling victim to scam calls, providing them with a sense of security and peace of mind.

2. LITERATURE SURVEY

One study [1] addressed this challenge by implementing a machine learning-based fraud detection system, focusing on logistic regression to achieve an accuracy of nearly 99%. Another research endeavor [2] explored the potential of Large Language Models, such as GPT-3.5 and GPT-4, for scam detection. The study highlighted the efficacy of these models in identifying various scams, including phishing and romance scams, while emphasizing the importance of continuous refinement in collaboration with cybersecurity experts to keep pace with evolving threats.In China, efforts to combat telecommunication fraud have involved leveraging machine learning and natural language processing[3]. By analyzing call content from diverse sources like news reports and social media, researchers have constructed datasets and devised rules for fraud detection. An Android application has been developed to facilitate real-time analysis of incoming calls, enabling the prevention of fraudulent activities and the safeguarding of customers.Credit card fraud detection is a critical concern in financial transactions. One study addresses this challenge by implementing a fraud detection system using machine learning, particularly focusing on logistic regression. The system, aiming for an accuracy of nearly 99%, enables administrators to authenticate, upload datasets, and identify fraudulent transactions via a web application. Using word embedding techniques in natural language processing, the Anti-Social Engineering Tool (ASsET) [4] identifies telephone scams by analysing semantic content and speech acts in conversations. Achieving high accuracy in distinguishing between scam and non- scam calls, ASsET contributes significantly to fraud prevention efforts.A novel method for detecting spam calls focuses on analysing acoustic features of incoming voicemails [5]. By distinguishing between human and robocalls with 93% accuracy and identifying spam calls with 83% accuracy, this approach complements existing strategies for mitigating unwanted calls and fraudulent activities.

Artificial intelligence plays a crucial role in detecting and analyzing fraudulent phone calls [6]. This approach utilizes real-world datasets to identify malicious calls and indicators of fraud, addressing concerns within the telecommunications industry effectively. The vulnerability to phone call-based fraud prompts research efforts, such as a scam detection system utilizing machine learning and MFCC features [7]. Achieving high accuracy, this system focuses on the Indonesian language and aims to mitigate the risks associated with fraudulent phone calls.In another study, machine learning techniques, including natural language processing and deep learning, are employed for scam call detection [8]. Leveraging datasets of scam and non-scam calls, the research achieves an accuracy of 85.61% with the LSTM algorithm, demonstrating the effectiveness of NLP techniques in identifying fraudulent activities. Furthermore, a novel approach to spam and scam call detection involves conversational AI models trained on a dataset of such calls [9]. By enabling real-time transcription and analysis of unknown calls, this method helps in detecting suspicious content and reducing the occurrence of spam calls.Lastly, a trust-based mechanism is proposed for VoIP spam detection, utilizing call duration and direction [10]. With adjustable trust values based on calling behavior, this mechanism effectively distinguishes between spam calls and legitimate ones in realistic scenarios.

3. PROPOSED METHODOLOGY:

The research endeavors to construct a real-time scam detection system utilizing speech-to-text classification. The system's primary objective lies in converting speech inputs into text, subsequently employing this text as the basis for a classification model. The project involves several pivotal stages to realize its goals:

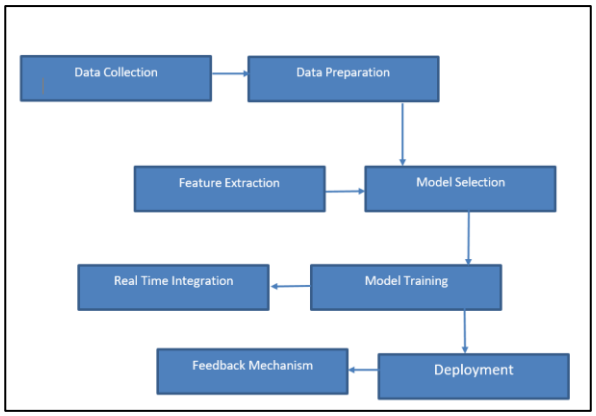

Figure 1. Proposed Methodology

1. Data Collection:

- A diverse dataset is compiled, encompassing audio recordings featuring both genuine and fraudulent conversations. The dataset's scope encompasses various scam types and scenarios, ensuring comprehensive training and evaluation of the detection system.

	First Word	Content
0	fraud	hello, i m bank manager of SBI, ur debit card ...
1	fraud	Todays Vodafone numbers ending with 4882 are s...
2	normal	Please don't say like that. Hi hi hi
3	normal	Thank you!
4	normal	Oh that was a forwarded message. I thought you...

Figure 2: Input data for training the model

2. Preprocessing:

- The collected audio files undergo preprocessing, where speech-to-text conversion techniques are applied to transcribe the audio content into textual format. Techniques such as noise reduction and standardization to lowercase are employed to cleanse and homogenize the text data.

- Tokenization is the process of dividing a text into discrete words or units.
- Eliminating frequent terms that don't help differentiate the courses is known as stop word removal.
- Lemmatization/stemming: reducing words to their simplest or root form.
- Vectorization: Using methods like TF-IDF (Term Frequency-Inverse Document Frequency), text is transformed into numerical characteristics.

3. Feature Extraction:

- From the preprocessed textual data, pertinent features are extracted to capture linguistic patterns indicative of fraudulent content. This involves deriving word frequencies, n-grams, sentiment scores, and potentially deploying advanced NLP techniques like word embedding's or pre-trained language models. TF-IDF vectorization was used to convert the text into a numerical representation. Each call transcript was transformed into a TF-IDF vector,

capturing the importance of words in the context of the entire dataset.

4. Model Selection:
   - The researchers explore a spectrum of classification models suited for detecting fraudulent content based on the extracted features. Machine learning algorithms such as Support Vector Machines, Random Forest, Gradient Boosting, and deep learning models like Recurrent Neural Networks and Transformer-based architectures are under consideration for model selection.

5. Model Training:
   - The chosen classification model undergoes training using the preprocessed and feature-extracted data. Parameters are fine-tuned and model performance is assessed using standard metrics like accuracy, precision, recall, and F1-score.

6. Real-time Integration:
   - A real-time speech-to-text pipeline is developed to continuously convert incoming audio streams into textual format. The text classification model is integrated into the pipeline to classify converted text as either genuine or fraudulent content in real-time.

7. Feedback Mechanism:
   - To refine the model's performance over time, a feedback mechanism is instituted to gather input from users or domain experts. This feedback is utilized to periodically retrain the model, enabling adaptation to evolving scam patterns and tactics.

8. Deployment:
   - The completed real-time scam detection system is deployed in a production environment with emphasis on scalability, reliability, and security considerations. Continuous monitoring of system performance and user feedback facilitates identification and resolution of any encountered issues or required enhancements.

9. User Interface (UI):
   - A user-friendly interface is devised to present real- time scam detection results. Users receive alerts or notifications upon detection of potential fraudulent content.

10. Compliance and Privacy:
    - Stringent measures are implemented to ensure adherence to pertinent regulations and privacy policies governing the processing and handling of sensitive audio and text data. Robust security protocols are instituted to safeguard user data and system integrity.

Through these methodically executed steps, the research aspires to contribute to the development of an effective and dependable real-time scam detection system leveraging speech-to-text classification.

4. IMPLEMENTATION:

The SpeechRecognition library to capture audio from the microphone and convert it to text.Here is the python code :

```
def speech_to_text():
print("Adjusting for ambient noise, please wait...")
recognizer.adjust_for_ambient_noise(source, duration=5)
print("Listening...")
audio_data = recognizer.listen(source) print("Recognizing...")
```

4.1 Model Training Using NLP AND SVM.

The process of training a model using Natural Language Processing (NLP) and Support Vector Machines (SVM) involves several key steps. Firstly, the data must be prepared by collecting and preprocessing textual information. This includes tasks such as tokenization, lowercasing, removal of stop words, and potentially stemming or lemmatization to standardize the text data.

Following data preparation, the dataset is typically divided into training and testing sets to facilitate model evaluation. The next crucial step is feature extraction, where the textual data is transformed into numerical feature vectors suitable for machine learning algorithms like SVM. Common techniques for feature extraction in NLP include Bag of Words (BoW), Term Frequency- Inverse Document Frequency (TF-IDF), and word embeddings.

Subsequently, an SVM classifier is trained using the preprocessed and transformed text data. SVM aims to find the optimal hyperplane that best separates the data points belonging to different classes, based on the extracted features.

Once the model is trained, it is evaluated using the testing set to assess its performance. Evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix are commonly used to measure the model's effectiveness in classification tasks.

Parameter tuning may also be performed to optimize the SVM model's performance. Parameters such as the regularization parameter (C) and the choice of kernel function (e.g., linear, polynomial, radial basis function) can significantly impact the model's accuracy and generalization capability.

Upon achieving satisfactory performance, the trained model can be deployed to make predictions on new, unseen data, thereby fulfilling the intended application of the NLP-SVM model in real-world scenarios.

This structured approach to model training using NLP and SVM ensures robustness and efficiency in addressing various text classification tasks while adhering to the principles of machine learning and natural language processing.

4.2 Naive Bayes

Naive Bayes is a widely used classification algorithm in machine learning, particularly suited for tasks like text classification, spam filtering, and sentiment analysis. It's based on Bayes' theorem, which calculates the probability of a hypothesis given the evidence.

What makes Naive Bayes "naive" is its assumption of feature independence. This means it treats each feature as independent of the others, which simplifies the calculations. In practice, this assumption might not hold true, but Naive Bayes often performs well regardless.

In text classification, Naive Bayes analyzes the occurrence of words or features in a document to determine its class.

For example, in spam filtering, it calculates the likelihood of an email being spam or not based on the words it contains.

One of the main advantages of Naive Bayes is its simplicity and efficiency. It's computationally inexpensive, making it suitable for large datasets. Moreover, it requires a relatively small amount of training data to estimate parameters, which is beneficial when working with limited resources.

However, Naive Bayes has limitations. Because of its independence assumption, it might not capture complex relationships between features accurately. Also, it's sensitive to features that are not present in the training data, leading to potential bias in predictions.

In research papers, Naive Bayes is often discussed as a baseline model for comparison with more complex algorithms. Its simplicity and reasonable performance make it a valuable tool for various classification tasks, especially in scenarios where interpretability and computational efficiency are crucial. Nonetheless, researchers should be mindful of its assumptions and limitations when applying Naive Bayes in their studies.

5. CONCLUSION

In conclusion, the development of a scam call tone analyzer with real-time data using NLP and machine learning algorithm (Naive Bayes) offers a promising solution to the growing issue of scam calls. This work aims to address this challenge by leveraging machine learning algorithm for the detection and analysis of such calls. By analysing the tone and content of incoming calls, this system can identify potential scams and terminate them under the user's permission. The integration of NLP and machine learning ensures accurate classification and reduces false positives, providing users with an effective defense against fraudulent activities over the phone.

REFERENCES

[1] Credit Card Fraud Detection Using Logistic Regression, Ameer Saleh Hussein, Rihab Salah Khairy, Shaima Miqdad Mohamed Najeeb, Haider Th. Salim ALRikabi, https://doi.org/10.3991/ijim.v15i05.17173

[2] Liming Jiang. 2024. Detecting Scams Using Large Language Models. 1, 1 (February 2024), 8 pages. https://doi.org/10.48550/arXiv.2402.03147

[3] Detecting telecommunication fraud by understanding the contents of a call, Qianqian Zhao, Kai Chen, Tongxin Li, Yi Yang and XiaoFeng Wang,31 August 2018, https://doi.org/10.1186/s42400-018-0008-5

[4] Detecting Telephone-based Social Engineering Attacks using Scam Signatures, Ali Derakhshan, Ian G. Harris, Mitra Behzadi, 26 April 2021, https://doi.org/10.1145/3445970.3451152

[5] Detection of Robocall and Spam Calls using Acoustic Features of Incoming Voicemails, Benjamin Elizalde, Dimitra Emmanouilidou, FEBRUARY 23 2022, https://doi.org/10.1121/2.0001533

[6] Detection and Analysis of Fraud Phone Calls using Artificial Intelligence, Saloni Malhotra, Ginni Arora, Ruchika Bathla

[7] Phone Call Speaker Classification using Machine Learning on MFCC Features for Scam Detection, Yanisa Medrina Rahman, Yoanes Bandung

[8] Scam Calls Detection Using Machine Learning Approaches, Brendan Hong, Tee Connie, Michael Kah Ong Goh

[9] Kolati Mallikarjuna and Patel, Bhavik Kumar, "Suspicious Call Detection and Mitigation Using Conversational AI", Technical Disclosure Commons, (December 04, 2023) https://www.tdcommons.org/dpubs_series/6473

[10] Trust-based VoIP Spam Detection based on Calling Behaviors and Human Relationships, Noppawat Chaisamran, Takeshi Okuda, Suguru Yamaguchi, November 2, 2012