Artificial Intelligence-Based Detection of Fraudulent Phone Calls for Secure Mobile Communication

Dr.B.V.S.T.Sai, Shaik Subhani V.Roshan Kumar

Professor in Dept of Mathematics and CSE, St.Mary's Group of Institutions Guntur, Email: bvstsai@gmail.com

Assoc Professor in Dept of CSE, St.Mary's Group of Institutions Guntur, Email: subbu.buddu@gmail.com

Assistant Professor in Dept of CS, Bapatla Engineering College, Email: roshan4linux@gmail.com

ABSTRACT: This study focuses on detecting mobile phone fraud calls using a hybrid approach that combines Support Vector Machine (SVM) and Region-based Convolutional Neural Networks (R-CNN). SVM is utilized for classifying call features such as duration, frequency, and user behaviour patterns, while R-CNN is applied to analyse mobile voice calls and extract relevant features from spectrogram images. By integrating SVM's efficiency in handling structured data with R-CNN's ability to process unstructured data like audio, the system achieves improved accuracy in identifying fraudulent calls. The model is designed to minimize false positives and enhance real-time detection, ensuring user security and privacy.

KEYWORDS: Malicious, Artificial Intelligence, Fraudulent, Machine learning.

1. INTRODUCTION

The rapid growth of telecommunications technology has led to a surge in fraudulent phone calls, commonly known as scam calls. These calls present serious threats, including financial fraud, identity theft, and data breaches. As scammers employ more sophisticated tactics, traditional detection and prevention methods struggle to keep up. Artificial Intelligence (AI) emerges as a powerful solution to enhance the detection and prevention of such malicious activities. Phone-based scams pose a persistent threat to individuals, businesses, and governments alike. In 2021, the U.S. Federal Trade Commission (FTC) received over 3 million fraud reports, resulting in total losses exceeding $3 billion. Scammers employ tactics such as impersonation, caller ID spoofing, and digital manipulation to steal sensitive information, money, or damage reputations. These fraudulent activities have global repercussions, causing significant financial and informational harm. AI-driven systems, leveraging machine learning, natural language processing, and behavioral analytics, represent a crucial advancement in combating phone fraud. As technology evolves, AI will play an increasingly essential role in protecting individuals and

organizations from the growing risks posed by scam calls.

The key contributions of our work are summarized as follows:

- We design and construct a classifier based on Calling Detail Records (CDR) for fraudulent phone call recognition. The classifier only uses the CDR as input data, so it can be constructed easily, quickly, and efficiently. It provides a basic framework for recognition task and defines the main steps of the task
- Our study provides the first systematic exploration of state-of-the-art machine learning algorithms applied to fraudulent phone call recognition, namely, We design, tune, and evaluate three models—the (RNN), SVM Our ML models are capable of automatically learning phone number features and call behaviour features for fraudulent phone call recognition. We demonstrate that our ML-based approach achieves a higher accuracy rate than the state-of-the-art approaches.
- We revaluate previous work on our new real-world datasets. As a result of a systematic comparison of our novel ML-based approach to previous fraudulent phone call recognition approaches, we demonstrate comparable recognition results with slight improvements of up to 3.0%-4.7% on average. Furthermore, our ML models reveal more general and stable phone number features and call behavior features of fraudulent phone calls than the state-of-the-art approaches, which make them more robust to concept drift caused by a highly dynamic fraudulent phone number and its call behaviour.
- We make the generated dataset publicly available, allowing researchers to replicate our results and systematically evaluate new approaches to fraudulent phone call recognition.

2. LITERATURE SURVEY

Fawcett and Provost [1] explored adaptive fraud detection methods in their seminal work published in *Data Mining and Knowledge Discovery*. They focus on enhancing fraud detection techniques by adapting to changing patterns in data. The authors discuss various algorithms and models used in fraud detection and emphasize the importance of dynamic

adaptation to new fraudulent strategies. Their research provides foundational insights into the development of adaptive systems capable of identifying and mitigating fraud effectively.

Weng et al. [2] addressed online e-commerce fraud through a large-scale detection and analysis study presented at the 2018 IEEE 34th International Conference on Data Engineering (ICDE). The paper highlights advanced detection methods for e-commerce fraud, leveraging large datasets and sophisticated analysis techniques. The authors discuss their approach's scalability and effectiveness in identifying fraudulent activities in online transactions, providing a comprehensive overview of current practices and challenges in e-commerce fraud detection.

3. SYSTEM ANALYSIS

3.1EXISTING SYSTEM

The existing system for "Detection and Analysis of Fraud Phone Calls using Artificial Intelligence" primarily relies on traditional methods and rule-based systems, often struggling to keep pace with evolving fraud tactics. Conventional call filtering techniques exhibit limitations in accurately distinguishing between legitimate and fraudulent calls. The absence of advanced machine learning models are This study presents a method for detecting fraudulent phone calls using Support Vector Machine (SVM), a supervised machine learning algorithm known for its effectiveness in binary classification tasks. The process involves collecting and pre-processing call data, including features such as call duration, frequency, and user behaviour patterns. SVM is then trained on labelled data to differentiate between legitimate and fraudulent calls. The model's ability to find optimal decision boundaries makes it suitable for identifying complex patterns associated with scam calls. This approach enhances the accuracy and efficiency of fraud detection, offering a scalable solution for real-time protection against telecommunication scams.

3.2 LIMITATIONS OF EXISTING SYSTEM

- The current system isn't very flexible when it comes to new and advanced fraud schemes.
- Target classes that overlap, or when there is a lot of noise in the data set, make SVM underperform.
- SVM performance will become degraded when there are more features per data point than there are samples of training data.

3.3 PROPOSED SYSTEM

The proposed system for "Detection and Analysis of Fraud Phone Calls using Artificial Intelligence" represents a paradigm shift, leveraging advanced machine learning algorithms for dynamic adaptation to evolving fraud tactics. This solution incorporates a comprehensive learning model, continuously evolving through real-time data analysis to identify emerging patterns of fraudulent behaviour. Additionally, The current step of a Recurrent Neural Network (RNN) takes as input the output of the previous step. Every input and output of a classic neural network do not rely on any other. Still, remembering the prior words is necessary while trying to guess the next word in a phrase since the previous words are needed. That is why RNN was born; it used a Hidden Layer to fix the problem. The Hidden state, which stores certain sequence-related information, is the most crucial and central component of RNN.

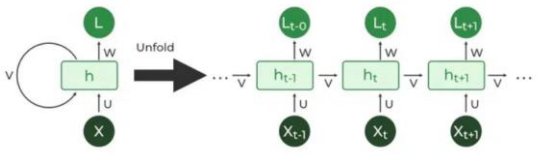

Fig: R-CNN Architecture

The proposed system prioritizes user feedback and employs a feedback loop mechanism to further refine its accuracy and reduce false alarms. offering heightened accuracy, adaptability, and insights into evolving fraud strategies in the realm of phone calls.

4. SYSTEM ARCHITECTURE

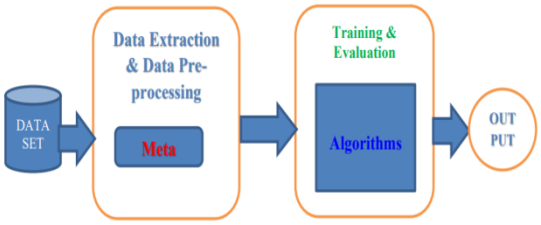

Fig 4: System Architecture

The architecture relies on a comprehensive dataset and metadata for analysis. Machine learning algorithms are deployed to detect patterns in the data. The output provides real-time insights into the authenticity of phone calls.

4.1 METHODOLOGY

- First remove duplicate data and missing values from the set.
- Transform categorical features such as call type, caller ID into numerical features. Using label encoding.
- Normalization of the numerical features, such as call duration, frequency, is done by using z-score normalization.

- Selection of suitable artificial intelligence-based model such as RNN, support vector machine (SVM), etc. Implement the model and train the pre-processed data. Here, RNN is selected for training and testing of the dataset for better prediction .

5. MODULES

i) Data collection: The dataset of phone call recordings, along with metadata such as call duration, location, phone numbers, etc. is collected from real world source like. Dataset is taken from real world sources such as Kaggle. The dataset contains 1000 genuine and fraudulent calls shown in fig 6.1 and the following features such as State, area code, a phone number, date and time, IP address, code, etc. this dataset is divided into two parts training set and testing set.

ii)Data Pre-processing: Data cleaning and pre-processing is used for dataset to remove noise, distortions, and irrelevant information.

iii) Model Evaluation: Selection of suitable artificial intelligence-based model such as RNN, support vector machine (SVM), decision tree, etc. Implement the model and train the pre processed data. Here, support vector machine and recurrent neural network is selected for training and testing of the dataset.

iv) User Interface (UI) Development: Constructs an intuitive and interactive user interface (UI) for CADM, facilitating user interaction and visualization of results generated algorithms.

6. RESULT

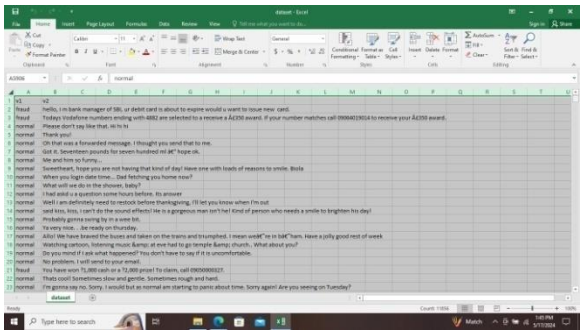

Fig 6.1: Phone Call database from Kaggle

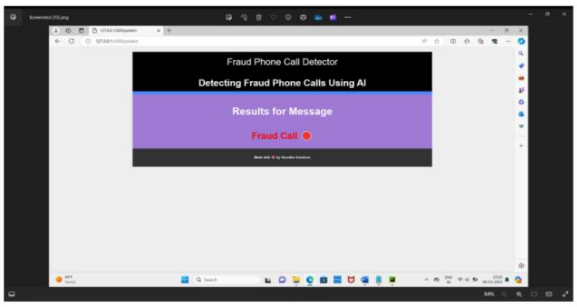

Fig 6.2: Phone Call Prediction by using Machine Learning

The proposed approach was evaluated on a dataset of 1000 genuine and fraudulent calls. the approach achieved a high accuracy of 90% and precision of 93.79%, outperforming the existing approaches. Hence, the approach was able to detect

fraudulent phone calls brilliantly. Below figures shows the results expected and actual result.

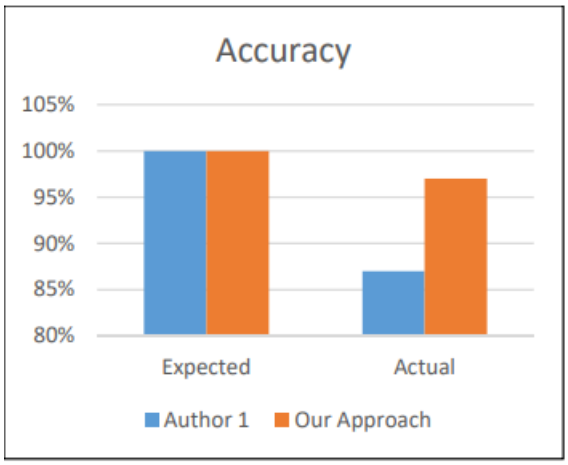

Fig 6.3: Accuracy of Fraudulent Calls Detection

In above, fig 6.3 it shows the accuracy of the approach applied by other author and the approach applied in this paper. The figure is shown in the form of expected result versus actual result. The expected accuracy is 100% and actual accuracy is 87% for author 1, whereas, the expected accuracy is 100% and actual accuracy is 97% for our approach. Hence, our approach is better.

7. CONCLUSION & FUTURE SCOPE

Fraudulent phone calls are a growing concern that affects individuals as well as organizations worldwide. The main purpose of this paper is to detect and analyze fraud phone calls using artificial intelligence. For achieving this goal, RNN, support vector machine (SVM), algorithms are used. The approach achieved a high accuracy and precision. In future To achieved a high accuracy and precision Implementing block chain for storing call records can provide a tamper-proof and transparent system for tracking calls, thereby reducing the chances of fraud. Hence, it will be a good solution to detect and analyze fraud or malicious calls.

REFERENCES

[1] T. Fawcett and F. Provost, "Adaptive Fraud Detection," Data Mining and Knowledge Discovery, vol. 1, pp. 291-316, 1997.

[2] H. Weng et al., "Online E-Commerce Fraud: A Large-Scale Detection and Analysis," 2018 IEEE 34th International Conference on Data Engineering (ICDE), pp. 1435-1440, 2018. [Online] Available: https://ieeexplore.ieee.org/document/8462781.

[3] S. M. Gowri, G. Sharang Ramana, M. Sree Ranjani and T. Tharani," Detection of Telephony Spam and Scams using Recurrent Neural Network (RNN) Algorithm," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 1284-1288, doi: 10.1109/ICACCS51430.2021.9441982.

[4] Abidogun, Olusola Adeniyi. "Data mining, fraud detection and mobile telecommunications: call pattern analysis with unsupervised neural networks." PhD diss., University of the Western Cape, 2005.

[5] S. Sandhya, N. Karthikeyan, R. Sruthi "Machine learning method for detecting and analysis of fraud phone calls datasets" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

[6] Mohammad Iquebal Akhter, Dr. Mohammad Gulam Ahamad "Detecting Telecommunication fraud using neural networks through data mining" international Journal of Scientific & Engineering Research, Volume 3, Issue 3, March-2012.

[7] I. Murynets, M. Zabarankin, R. P. Jover and Panagia, "Analysis and detection of SIMbox fraud in mobility networks," IEEE INFOCOM 2014 - IEEE Conference on Computer Communications,Toronto,ON,Canada,2014, pp.1519-1526,doi: 10.1109/INFOCOM.2014.6848087.

[8] Crawford, M., Khoshgoftaar, T.M., Prusa,J.D. et al. Survey of review spam detection using machine learning techniques. Journal of Big Data 2, 23 (2015). doi:10.1186/s40537-015-0029- 9.

[9] Marzuoli A, Kingravi H, Dewey D and Pienta R. (2016). Uncovering the Landscape of Fraud and Spam in the Telephony Channel 2016 15th IEEE International Conference on Machine Learning and Applications(ICMLA).10.1109/ICMLA.2016.0 153. 978-1-5090-6167-9. (853- 858).

[10] B. Teh, M. B. Islam, N. Kumar, M. K. Islam and U. Eaganathan,"Statistical and Spending Behavior based Fraud Detection of Card-based Payment System,"2018 International SSConference on Electrical Engineering and Informatics (ICELTICs), Banda Aceh, Indonesia, 2018, pp. 78-83, doi:10.1109/ICELTICS.2018.8548878.

[11] H. Tu, A. Doupe, Z. Zhao, and G.-J. Ahn, " Sok: Everyone hates 'robocalls: A survey of techniques against telephone spam," 2016 IEEE Symposium on Security and Privacy (SP), pp. 320 338, 2016.

[12] M. Crawford, T.M. Khoshgoftaar, J.D Prusa, A.N. Richter, H. Al Najada, "Survey of review spam detection using machine learning techniques", Journal of Big Data, 2, pp. 1-24, 42015.