7th European Conference on Industrial Engineering and Operations Management
Augsburg, Germany, July 16-18, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/EU07.20240070
Published: July 16, 2024

A Novel Ensemble-based Machine Learning Model for Anomaly Detection in CDRs to Identify International Revenue Share Fraud

Remalia Mayeni, Sibusisiwe Dube and Belinda Ndlovu
Department of Data Analytics and Informatics, Faculty of Applied Sciences
National University of Science and Technology, Bulawayo, Zimbabwe
P.O Box AC939, Ascot, Bulawayo, Zimbabwe
remaliamayeni@gmail.com, sibusisiwe.dube@nust.ac.zw, belinda.ndlovu@nust.co.zw,

Martin Maduva
Department Information and Marketing Science Department
Midlands State University
Senga Road, Gweru, Zimbabwe
muduvam@staff.msu.ac.zw

Fungai Jacqueline Kiwa
Department of ICT and Electronics at Chinhoyi University of Technology
Chinhoyi University of Technology
Private Bag 7724, Chinhoyi
jkiwa@cut.ac.zw

Abstract

Mobile network operators in developing countries often rely on traditional fraud detection systems, overlooking the potential of advanced machine learning techniques. This study addresses this gap by developing an International Revenue Share Fraud (IRSF) detection model using ensemble learning with random forest and support vector machine algorithms. The model analyzes Call Detail Records (CDRs) to identify fraudulent call patterns. CDRs contain call attributes like time, duration, source and destination numbers and completion status, providing valuable data for anomaly detection. Random Forest is chosen for its effectiveness in handling complex and imbalanced datasets, common in telecom fraud scenarios. Its ability to address imbalanced data is crucial, as fraudulent calls are typically rare compared to legitimate ones. This research aims to develop a machine learning model that leverages call logs to detect fraudulent international account takeover. Our results advance descriptive analysis and improve knowledge of the traits and patterns of IRSFs. In the end, this produces a picture of IRSF operations that is more accurate. The model demonstrates good predictive performance on the testing set with a Mean Absolute Error (MAE) of 1.1208, indicating a low average absolute difference between predicted and actual values and the R-squared value of 0.9828 signifying strong overall predictive accuracy.

Keywords

Telecommunication, call detail record (CDR), fraud detection, machine learning, ensemble learning, Random Forest

1. Introduction

Telecommunications has become increasingly important in modern society, with customer needs no longer being limited to just connectivity, but also requiring access anywhere and anytime. This puts pressure on telecommunications operators and also increases the risk of new forms of digital fraud. In an industry where profit margins are increasingly tight, fraud represents the most significant avoidable threat to revenue for mobile operators, and in some regions, IRSF is a major

contributor to this issue Gent (2021).

Ensemble learning is a machine learning technique that aims to enhance predictive performance by combining predictions from multiple models. This method utilizes the strengths of different algorithms to improve the overall model performance (Mahajan et al. 2023). Anomaly detection is a crucial task in machine learning, focusing on identifying data instances that significantly deviate from expected or normal patterns. In the telecommunications sector, anomaly detection is commonly used to detect potential fraudulent activities or network irregularities within Call Detail Records (CDRs). This field has been extensively researched across various scientific disciplines and practical applications (Raghavendra Chalapathy and Chawla 2019).

In telecommunication networks (telecoms), a continuous stream of data on various parameters is generated such as signal strength and data usage and information reflecting subscriber activity. One of these metrics, call detail records (CDR), are frequently used to identify subscriber behavior patterns inside a network (Mokhtari et al. 2022). By analyzing these real-time and historical measurements, telecom companies can pinpoint areas where network performance is subpar. This allows them to take targeted actions such as optimizing cell tower configurations, adjusting network capacity allocation and even deploying new infrastructure in high-demand areas.

Normal Revenue Sharing: Telecom companies frequently enter into agreements to share a portion of the revenue generated from international calls. This is due to the fact that calls often traverse multiple networks to reach their final destination and each network involved receives a portion of the total cost. Interconnection charges are paid by one service provider to another for completing calls that originate from their network but terminate in another network.
International Revenue Share Fraud (IRSF) is a type of telecom fraud that exploits revenue-sharing agreements between telecommunication companies (Mohite 2023). It is when a fraudster makes an agreement with a local carrier in high cost destination to share profit for increasing traffic. The fraudster then gains unauthorized access to any organization's public branch exchange (PBX) and uses it to make calls (Ighneiwa and Mohamed 2017). The cost of making international phone calls is a significant driver of telecommunications fraud. Fraudsters can exploit the difference between high international call rates and lower local call rates for profit (Michel et al. 2023). This disparity is particularly concerning in developing economies, such as those in Africa, where international call termination rates remain comparatively elevated.

One prominent technique employed by telecommunication fraudsters is SIM box bypass fraud. This method involves the fraudulent rerouting of international calls over Voice over Internet Protocol (VoIP) networks. The calls are then terminated back onto the legitimate operator's network, masquerading as local traffic (Michel et al. 2023). This fraudulent manipulation allows perpetrators to exploit the disparity in call termination costs between international and local calls. While the per-call cost within large networks has decreased significantly, the aggregate financial losses incurred through this method can be substantial due to the bypass of foreign settlement charges and domestic access fees levied between telecommunication companies. Before a service provider's fraud management system becomes aware of the fraud, fraudsters can make big gains by driving massive volumes of traffic to the high-risk numbers swiftly iconectiv (2016). Scammers look for places to route calls where the fees for completing the call are very high. These places might also be offering a large portion of the call revenue back to the company that started the call. The research aims to develop a machine learning model that leverages call logs to detect fraudulent international account takeover providing valuable insights into the latest developments in the field of telecom fraud and helping stakeholders to stay abreast of emerging trends in this domain.

1.1 Objectives

The objectives of this research are to:

- detect fraudulent call activities on the network towards the international market.
- create a model that predict fraudulent and non-fraudulent calls.

The paper follows a structured format. Section 1 provides the Introduction giving the background context. Section 2 contains a thorough literature review. Section 3 outlines the methodologies utilized. Section 4 describes data collection. Section 5 presents results and validation. The conclusion is in the final section.

2. Literature Review

By building on earlier findings, literature reviews and evidence syntheses are crucial exploratory outputs that aid in the gradual advancement of research (Lame 2019). Mokhtari et al. (2022) developed a model grounded on call detail records (CDRs) from mobile network drivers to understand user behavior and identify anomalies in network traffic. The research they conducted offered a new hybrid approach that combines techniques for anomaly identification. K-means clustering which is a machine learning method that groups comparable data points. and neural networks which are a type of artificial intelligence that can learn complicated patterns from data. Ni and Yu (2022) concurred and conducted a research aimed

to develop a new model to assess the consequences of telecom fraud on victims, using Bayesian networks (BN). This model would assist law enforcement and community organizations in fraud prevention. The model was built by analyzing real-world telecom fraud cases and expert knowledge but there were some limitations that is the model's effectiveness relied heavily on the quality and quantity of historical fraud case data and the model assumed experts providing knowledge have equal weight, which might not always be true.

Several recent studies, including those by Michel et al. (2023) and Daka (2022), have addressed the issue of SIM box fraud detection and prevention in the telecommunications industry, albeit using different methodologies and datasets. In their research, Michel et al. (2023) evaluated the efficacy of three supervised machine learning algorithms in identifying fraud in Call Detail Record (CDR) data that was received from a network: Random Forest, Support Vector Machine (SVM), and XGBoost. Their findings indicated that Random Forest outperformed the other algorithms, achieving 92% accuracy in detecting SIM box fraud. The authors stressed the significance of real-time CDR analysis in their approach. On the other hand, Daka (2022) focused on a single supervised learning method: Artificial Neural Networks (ANN) with a Multi-layer perceptron (MLP) classifier. The model demonstrated 100% accuracy on the test dataset, potentially due to the balanced nature of the data. The study underscored the importance of analyzing call traffic patterns to identify fraudulent activities and suggested that ANNs hold promise for SIM box fraud detection, as they can detect unusual calling patterns associated with fraud.

Ni and Yu (2022), along with Trinh et al. (2019), have conducted research to develop models that can help prevent future issues in telecommunication networks. This includes detecting abnormal traffic patterns and predicting individuals' vulnerability to fraud.Trinh et al. (2019) developed a semi-supervised deep learning framework using Long Short-Term Memory (LSTM) neural networks to identify anomalies. They trained their algorithms on periods with regular traffic patterns and tested them on known events such as football matches and flea markets. Their methods show high precision in detecting these events by identifying deviations from established traffic patterns. On the other hand, Ni and Yu (2022) focused on analyzing and predicting telecom fraud using a Bayesian Network (BN) model. The authors constructed this model from the perspective of the victim, considering factors like age, gender, occupation and knowledge of telecom fraud. The model utilized data from 533 real telecom fraud cases to understand the probability distributions of different fraud scenarios and their potential outcomes, including the type of fraud, the liability of reporting it and implicit fiscal loss.

Sahin and Francillon (2021) studied International Revenue Share Fraud (IRSF) using three methods: Firstly, a honeypot that mimicked unused phone numbers to see if they would be targeted by fraudsters. Followed by a large dataset of test calls and international premium rate numbers (IPRNs) collected over four years. Then Real call data from a European operator that they analyzed for IRSF. By comparing the honeypot and test data, the researchers identified features that could be used to identify IRSF in real-world calls. They tested their method on the real call data and achieved a 98% accuracy rate in detecting fraudulent calls with a very low false positive rate of 0.28%. Jabbar and Suharjito (2020) conducted research aimed to develop a method for detecting fraud in telecommunication Call Detail Records (CDRs) using machine learning. They used unsupervised machine learning algorithms, specifically K-Means and DBSCAN, to group similar calls together. The hypothesis is that these groups would have unique patterns that would identify phony calls. According to study by Rekha et al. (2021), researchers may find it difficult to draw conclusions from systems, make predictions, or make decisions for specific applications like fraud detection when dealing with class imbalanced data. To assess the accuracy of the models, Jabbar and Suharjito (2020) compared the machine learning groups to calls that were known to be false. In this dataset, they discovered that K-Means outperformed DBSCAN in terms of fraud detection.

A study was presented by Kassa et al. (2020) to enhance the precision of fraud unearthing in telecommunication data utilizing ensemble machine learning approaches. The research focused on this machine learning approach that combines multiple classifiers to achieve higher accuracy than individual ones. The study found that the ensemble method combining a J48 decision tree algorithm with 10-fold cross-validation achieved the best performance, reaching an accuracy of 96.73%. This suggests that ensemble methods can significantly improve fraud detection accuracy compared to using individual classifiers.

2.1 Research Gap from Existing Literature

Despite significant advancements in telecommunication fraud detection, there are still areas that require further investigation and exploration. A critical practical research gap persists in the limited body of studies dedicated to International Revenue Share Fraud (IRSF) detection within the African context. While some empirical investigations have been conducted in specific African countries, such as Zambia, Daka, (2022) the overall research landscape remains relatively sparse. This paucity presents a significant challenge, as the African continent possesses unique characteristics that likely influence IRSF patterns and the efficacy of detection methods. These unique factors include diverse calling

patterns, regulatory frameworks distinct from other regions and potentially limited data availability. To address this gap and enhance IRSF detection efforts in Africa, future research should prioritize comprehensive investigations specifically tailored to this context. Such endeavors should aim to develop and test IRSF detection methodologies that account for the aforementioned contextual factors, potentially including strategies to overcome data limitations and identify IRSF patterns relevant to African telecommunication behaviors.

A methodological challenge in machine learning arises from imbalanced class distributions within datasets (Rekha et al. 2021). In the context of IRSF detection, this translates to a scenario where fraudulent calls represent a significantly smaller proportion of data compared to legitimate calls. This class imbalance can lead to biases in machine learning models, where the model prioritizes identifying the majority class (legitimate calls) and performs poorly in detecting the minority class (fraudulent calls). To overcome this challenge and improve IRSF detection accuracy, future research should prioritize the development and application of techniques that effectively handle imbalanced data. Promising avenues include exploring oversampling approaches to increase the representation of the minority class, under sampling techniques to reduce the majority class size and investigating the use of ensemble learning methods that are more robust to class imbalances.

3. Methodology

In this research, the fraud detection model, specifically, the Sim box form of fraud was implemented using the CRISP-DM process model. Numerous data miners have endorsed the CRISP-DM model as the most effective model given that it promotes appropriate procedures and provides institutions with the framework needed to realize better, faster outcomes from data mining (Schröer et al. 2021). Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment are its six iterative phases.

The telecommunications industry generates an enormous amount of data, making manual analysis nearly impossible (Mokhtari et al. 2022). Telecom fraud is a growing problem globally, causing significant revenue loss for operators. This research provides an overview to help telecom executives invest in fraud protection solutions that can effectively and proactively safeguard their networks. Companies are increasingly using data-driven methods to detect fraud, particularly in international revenue share fraud detection. Machine learning models are trained using historical call detail record (CDR) data to recognize patterns and adjust security rules to prevent future fraud attempts (Baesens et al. 2021). The data preparation phase involved exploring and preprocessing the dataset, aligning with the research methodology's Exploratory Data Analysis (EDA) and feature selection steps. This was crucial for gaining insights into the CDR characteristics and ensuring clean, consistent and model-ready data. The CDR data was used to test the model's effectiveness in detecting telecom fraud calls. Experiments were designed to compare the performance of the proposed algorithm with traditional classification algorithms. The results were analyzed to determine if the fraud detection model meets the business success criteria. This critical stage informs decisions about deployment, further improvements, or the development of a new model. The goal was to implement the machine learning model in a telecommunication network by integrating it into the existing infrastructure and ensuring seamless compatibility with the operational environment.

4. Data Collection

The research utilized a telecommunications Call Detail Record (CDR) dataset procured from the risk department of a well-established local telecommunications provider. This dataset specifically targeted transactions associated with the Interconnection Border Control Function (IBCF) and the Session Initiation Protocol (SIP). The data comprised of fields, each record designed to provide details pertaining to individual calls as shown in Table 1. Collectively, these attributes facilitated a comprehensive comprehension of the telecommunications environment, with a particular emphasis on IBCF and SIP transactions. Subsequently, feature engineering techniques were employed to identify and incorporate the most influential features that demonstrably contribute to the envisaged outcome.

Table 1. Attribute Description

Attribute	Description
List-Of-Calling-Party-Address	Source of the call
Called-Party-Address	Destination of the call
Service-Delivery-Start-Time-Stamp	Call session begin
Service-Delivery-End-Time-Stamp	Call session end
Duration	Duration of call
ibcfTrunkGroupID	Incoming and outgoing destination
Role-Of-Node	Node is originating or terminating session

5. Results

Jupyter Notebook was employed as the IDE for coding, running and establishing the model, offering an interactive environment for data exploration and model development. Prior to the analysis, several packages such as pandas, numpy and matplotlib libraries were installed for data manipulation, analysis and visualization to uproot acumen from the CDR. Exploratory data analysis involved the use of functions like describe (), info () and shape () to gain an understanding of the dataset's characteristics and structure. These initial steps laid the groundwork for further data analysis and modeling.

As seen in Figure 1, the distribution plot of calls based on the IBCF node's role offers important insights into the type of calls that flow through the node. The IBCF serves as a gateway to external networks, providing NAT and Firewall functions and acts as the gateway to and from risky destinations where IRSF is most likely to occur. The aggregate findings show that the gateway switch is essentially acting as the originator for most calls (137821), followed by the terminating role (51551), with a smaller count for the back-to-back user agent (B2BUA) of (2193).

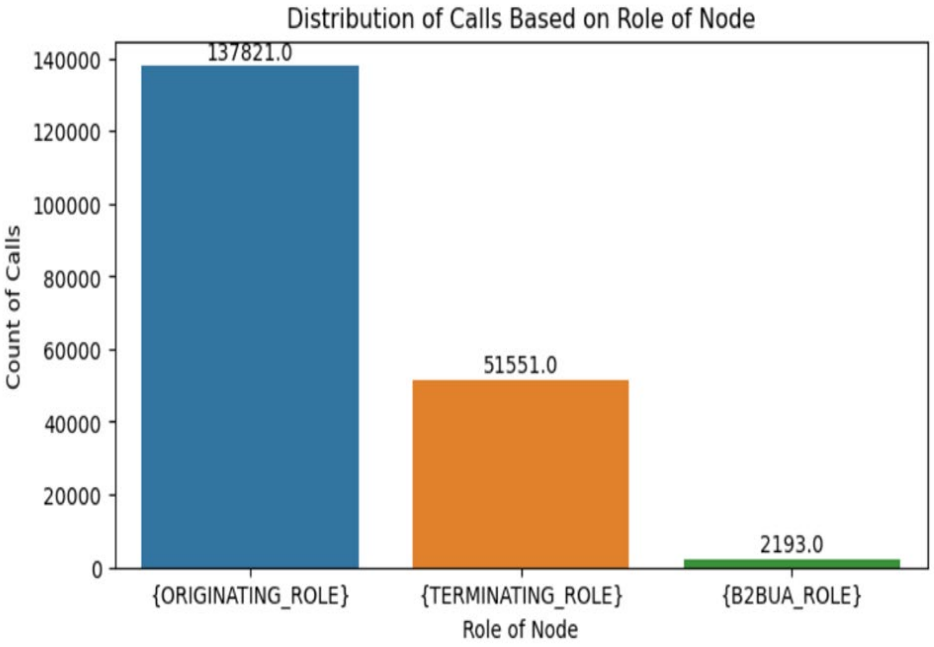

Figure 1. Distribution of calls on IBCF node

.

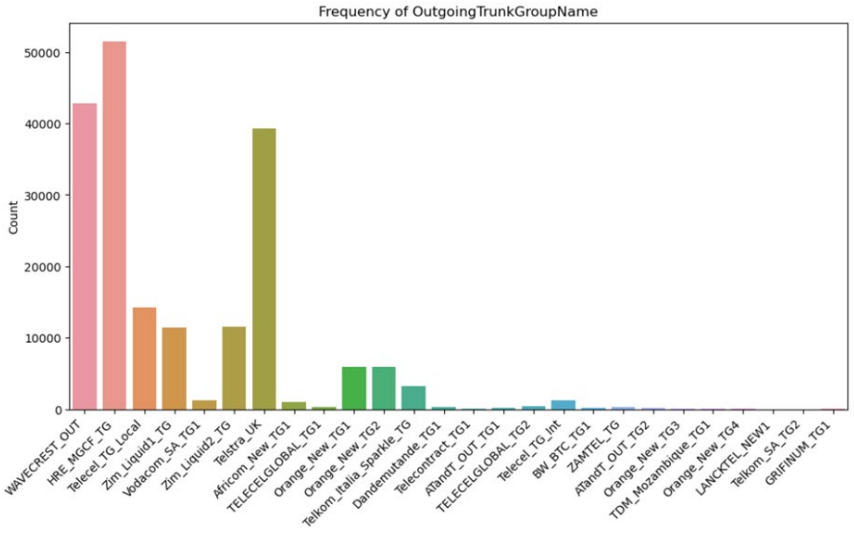

Figure 2. Frequency of Outgoing TrunkgroupName

The frequency of outgoing calls to international destinations was determined by monitoring traffic from multiple ISPs as shown in Figure 2 .This method was employed to identify fraudulent or high-risk activities on the network by tracking traffic flow towards the international market. International call routing involves several transit operators facilitating the connection between the originating and destination operators. Deliberate misrouting of calls by fraudulent telecom operators can lead to revenue losses for legitimate operators.

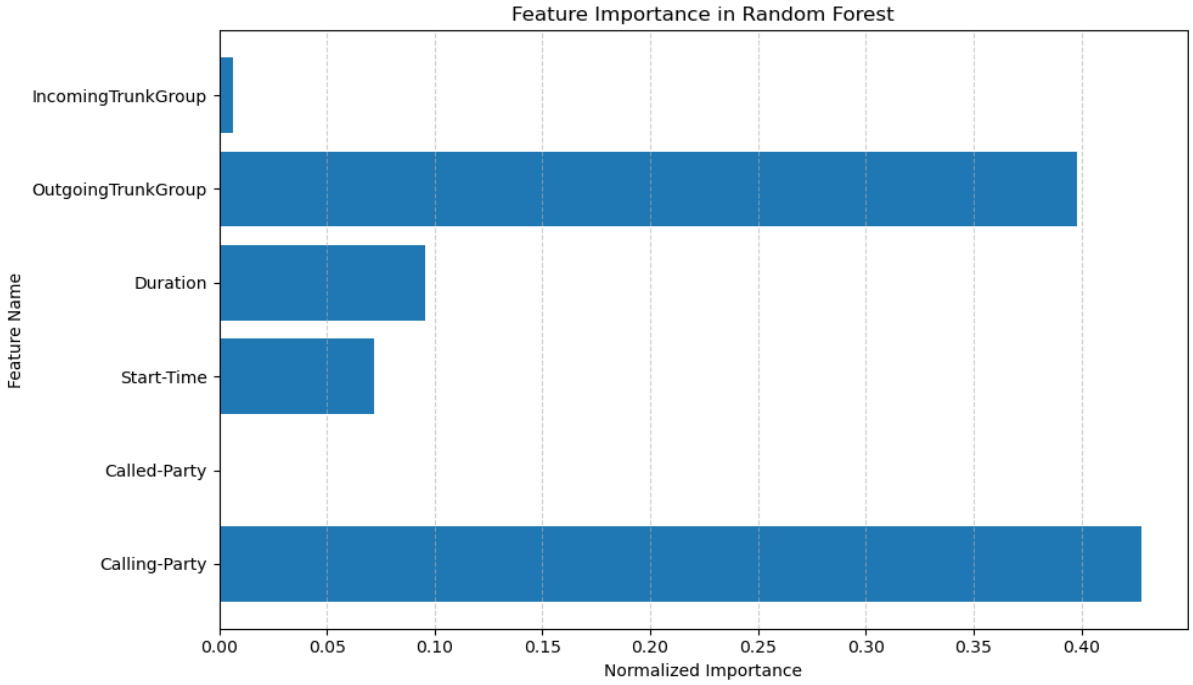

Figure 3. Feature Importance in Random Forest

In Figure 3 the feature importance plot generated using a Random Forest model is displayed. A prevalent ensemble learning approach Random Forest builds several decision trees and aggregates their forecasts to increase the model's overall robustness and accuracy. The most important feature indicated in the plot is "Calling-Party," suggesting that this feature has the most significant influence on the model's predictions. The Calling Party is where the call session is

generated and it's where the telecoms network bills. Most often, fraudsters will break into an organization's telephone network and start making calls to premium international revenue share numbers. The next most important feature is "OutgoingTrunkGroup," followed by "Duration," "Start-Time," "IncomingTrunkGroup," and "Called-Party."

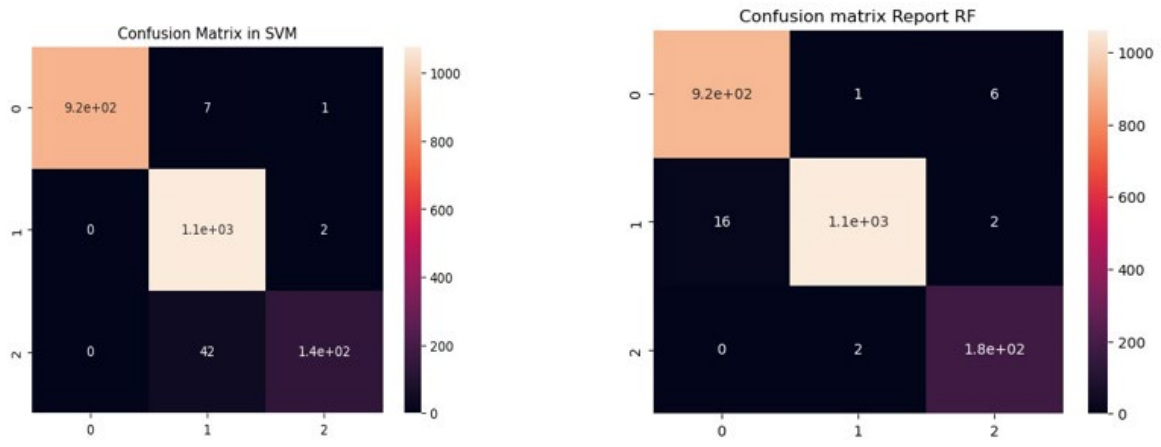

Figure 4. Confusion matrix for SVM and Random forest

The random forest and support vector machine methods were used to assess the fraud detection classification model's performance. The data was categorized into three classes: 0 for not fraud, 1 for suspicious and 2 for fraud. The confusion matrix in Figure 4 compared the actual data labels to the labels predicted by the model. SVM performed well in the non-fraud class compared to random forest, which misclassified 16 non-fraud calls as suspicious. For classes 1 and 2, random forest performed better than SVM.

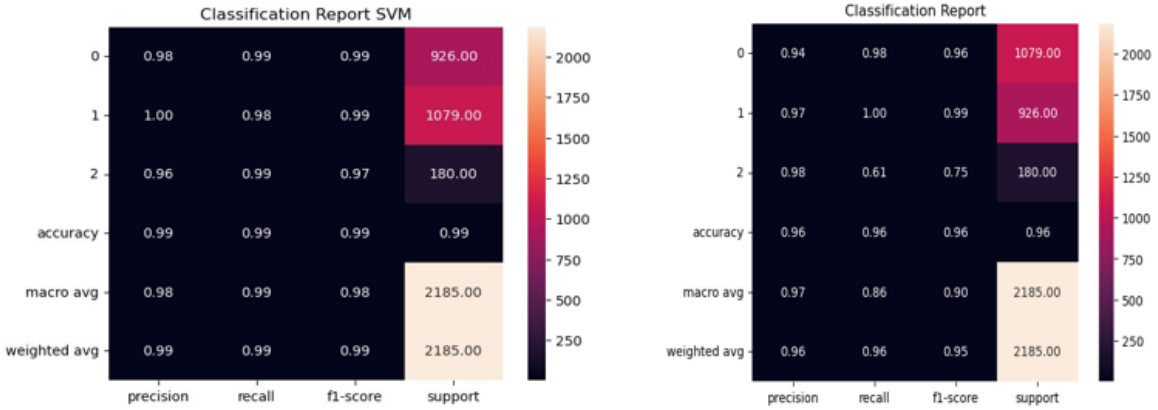

Figure 5. Classification reports for SVM and Random Forest

The overall weighted average accuracy, precision, recall and F1-score for both the SVM and Random Forest models are depicted in Figure 5. This indicates that overall, the models performed well on the testing data and were more effective at identifying fraudulent calls in the SVM model compared to the 0.96 accuracy of the Random Forest model. When looking at individual classes, the performance metrics vary. The first class of the Random Forest algorithm exhibits the highest precision (0.98), while the third class shows the highest recall (1.00) and F1-score (0.99).

5.1 Discussion

The effectiveness of a fraud detection system relies heavily on the accuracy, completeness, and suitability of the specifically calibrated data (Kasra Babaei et al. 2020). Data imbalance can be problematic, as certain types of fraudulent activities may be underrepresented in the data, making it difficult to accurately identify and detect these activities. The gateway switch from which the CDRs for this paper were extracted had a maximum limit of 16110 entities it could download at a time. To overcome this limitation, the researcher had to concatenate multiple entities to obtain a sufficient amount of records for model building. Building effective AI and ML models requires substantial amounts of data, posing a challenge in this particular scenario. The findings align with those of Jabbar and Suharjito (2020), who concluded that fraud issues continually evolve and their resolution varies with each case. Therefore, the outcomes of our research may

not necessarily yield favorable results if extrapolated to different telecommunications providers or varied instances of fraud.

5.2 Proposed Improvements

The research presented in this paper has made significant progress in evaluating the efficiency of the proposed fraud detection model. This paves the way for its potential implementation in real-world telecommunication systems. It's important to ensure seamless integration and compatibility with the existing telecommunication infrastructure to enable a smooth transition and adoption of the model. Additionally, the model's performance should be optimized for real-time processing to enable timely identification of fraudulent activities, minimizing latency and maintaining efficient integration with call processing systems. It's also significant to manage the scalability of the model to manipulate the voluminous quantity of data generated by telecommunication meshes, securing that the model can reuse considerable scale data without risking its delicacy or trustability.

5.2 Validation

The results of the model evaluation shed light on how efficiently the Random Forest algorithm performs on the training and testing sets (Table 2).The training set metric in Table 2 show the Mean Absolute Error (MAE) of 0.2674 indicates the average absolute difference between the actual and predicted values. A low MAE suggests that the model is relatively accurate in its predictions. The Mean Squared Error (MSE) of 1.0568 further quantifies the overall squared difference between actual and predicted values. The R-squared value of 0.9999 signifies an almost perfect fit of the model to the training data. These metrics collectively demonstrate the high accuracy and precision achieved during the training phase. The testing set metrics, the MAE of 1.1208 indicates the average absolute prediction error on the unseen data. While slightly higher than the training set MAE, it remains relatively low, implying good predictive performance. However, the MSE of 2.7873 is noticeably larger, reflecting increased variability in the prediction errors. On the testing set, the R-squared value of 0.9828 shows good predictive potential.

Table 2. Model results

Training Set Metrics: Mean Absolute Error: 0.26748382355859296 Mean Squared Error: 1.0568304844676841 R-squared: 0.999981575432612
Testing Set Metrics: Mean Absolute Error: 1.1207711823750357 Mean Squared Error: 2.7873454363405 R-squared: 0.9827781415660845

6. Conclusion

In conclusion, the development, design and assessment of the machine learning-based fraud detection model for telecommunication networks have yielded promising results in identifying fraudulent activities. This demonstrates its potential for real-world deployment. Another important objective of the study was to detect fraudulent call activities on the network specifically targeting the international market. The implemented Random Forest model has exhibited exceptional accuracy, evident from the evaluation metrics on both the training and testing sets. The low Mean Absolute Error and Mean Squared Error on the training set reflect the model's ability to closely approximate actual values. Although there is a minor increase in errors on the testing set, the high R-squared value indicates a robust correlation between the predicted and actual values. These findings strongly align with the goal of creating a predictive model capable of effectively identifying fraudulent events on the network.

The future of fraud detection will undoubtedly be transformed by AI and ML. These cutting-edge technologies will continuously progress, enabling increasingly effective identification and prevention of fraud. Telecommunication service providers are strongly advised to integrate and fully embrace these models into their systems to combat fraudulent activities and optimize revenue generation. These advanced models can swiftly and precisely analyze vast amounts of data, decisively reducing and proactively combating the occurrence of IRSF. Moreover, they can adeptly adapt to new forms of fraud, empowering organizations to consistently stay ahead of criminal activities.

References

Baesens, B., Höppner, S. and Verdonck, T. Data Engineering for Fraud Detection. *Decision Support Systems*, 150,.

doi:https://doi.org/10.1016/j.dss.2021.113492, p.113492, 2021.

Gent, A. SIM boxing – why mobile regulators need to act now. *Computer Fraud & Security*, 2021(5), doi:https://doi.org/10.1016/s1361-3723(21)00052-x, pp.9–12, 2021.

Getahun, A. and To, K. *Telecom Voice Traffic Termination Fraud Detection Using Ensemble Learning: The Case of Ethio Telecom*. [online] *http://repository.smuc.edu.et/*. Available at: http://hdl.handle.net/123456789/6226 [Accessed 22 Jun. 2024], 2020.

Iconectiv. *stop telecom fraudsters in their tracks*. [online] Available at: https://iconectiv.com/sites/default/files/2016-12/mobileid_wp_stop-telecom-fraudsters.pdf [Accessed 26 Jun. 2024], 2016.

Ighneiwa, I. and Mohamed, H. *Bypass Fraud Detection: Artificial Intelligence Approach Bypass Fraud Detection: Artificial Intelligence Approach*. [online] Available at: https://www.researchgate.net/publication/321070233_Bypass_Fraud_Detection_Artificial_Intelligence_Approach/citation/download [Accessed 22 Jun. 2024], 2017.

Jabbar, M.A. and Suharjito, S. Fraud Detection Call Detail Record Using Machine Learning in Telecommunications Company. *Advances in Science, Technology and Engineering Systems Journal*, [online] 5(4), doi:https://doi.org/10.25046/aj050409, pp.63–69, 2020.

Kasra Babaei, Chen, Z. and Maul, T. A Study of Fraud Types, Challenges and Detection Approaches in Telecommunication. *Journal of Information Systems and Telecommunication (JIST)*, 4(28), doi:https://doi.org/10.7508/jist.2019.04.002, p.248, 2020.

Lacuška, M. and Peráček, T. *Trends in global telecommunication fraud and its impact on business*.. Developments in Information & Knowledge Management for Business Applications: Volume 1. pp.459–485, 2021

Lame, G. Systematic Literature Reviews: an Introduction. *Proceedings of the Design Society: International Conference on Engineering Design*, 1(1), doi:https://doi.org/10.1017/dsi.2019.169. pp.1633–1642, 2019.

Mahajan, P., Uddin, S., Hajati, F. and Moni, M.A. Ensemble Learning for Disease Prediction: A Review. *Healthcare*, 11(12), doi:https://doi.org/10.3390/healthcare11121808, p.1808, 2023.

Michel, E., Kabiena Ivan Basile, Tchapga Tchito Christian, Ferry and Michael Ekonde Sone., Machine Learning-Based Approach for Identification of SIM Box Bypass Fraud in a Telecom Network Based on CDR Analysis: Case of a Fixed and Mobile Operator in Cameroon. *Journal of computer and communications*, 11(02), pp.142–157. doi:https://doi.org/10.4236/jcc.2023.112010, 2023.

Mohite, J. *Understanding International Revenue Share Fraud*. [online] https://www.akamai.com/. Available at: https://www.akamai.com/blog/security/understanding-international-revenue-share-fraud [Accessed 25 Jun. 2024], 2023.

Mokhtari, A., Ghorbani, N. and Bahrak, B. Aggregated Traffic Anomaly Detection Using Time Series Forecasting on Call Detail Records. *Security and Communication Networks*, [online] 2022, p.e1182315. doi:https://doi.org/10.1155/2022/1182315, 2022.

Ni, P. and Yu, W., A Victim-Based Framework for Telecom Fraud Analysis: A Bayesian Network Model. *Computational Intelligence and Neuroscience*, [online] 2022, p.e7937355. doi:https://doi.org/10.1155/2022/7937355, 2022.

Nyirenda, M. and Daka, J. Smart Mobile Telecommunication Network Fraud Detection System Using Call Traffic Pattern Analysis and Artificial Neural Network Smart Mobile Telecommunication Network Fraud Detection System Using Call Traffic Pattern Analysis and Artificial Neural Network. *American Journal of Intelligent Systems*, 2022(2), doi:https://doi.org/10.5923/j.ajis.20221202.01, pp.43–50, 2023.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S. and McGuinness, L.A. The PRISMA 2020 statement: an Updated Guideline for Reporting Systematic Reviews. *British Medical Journal*, 372(71). doi:https://doi.org/10.1136/bmj.n71, 2021.

Raghavendra Chalapathy and Chawla, S., Deep Learning for Anomaly Detection: A Survey. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.1901.03407, 2019.

Rekha, G., Tyagi, A.K., Sreenath, N. and Mishra, S., Class Imbalanced Data: Open Issues and Future Research Directions. *2021 International Conference on Computer Communication and Informatics (ICCCI)*. doi:https://doi.org/10.1109/iccci50826.2021.9402272, 2021.

Sahin, M. and Francillon, A., Understanding and Detecting International Revenue Share Fraud. *Proceedings 2021 Network and Distributed System Security Symposium*. [online] doi:https://doi.org/10.14722/ndss.2021.24051, 2021.

Schröer, C., Kruse, F. and Marx Gómez, J., ScienceDirect A Systematic Literature Review on Applying CRISP-DM Process Model-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/) Peer-review under responsibility of the scientific committee of the CENTERIS -International Conference on ENTERprise Information Systems / ProjMAN -International Conference on Project MANagement / HCist -International Conference on Health and Social Care Information Systems and Technologies ScienceDirect A Systematic Literature Review on Applying CRISP-DM Process Model-NC-ND license

(http://creativecommons.org/licenses/by-nc-nd/4.0/) Peer-review under responsibility of the scientific committee of the CENTERIS -International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist -International Conference on Health and Social Care Informatio…. *Procedia Computer Science*, 181, doi:https://doi.org/10.1016/j.procs.2021.01.199, pp.526–534, 2021.

Tu, K., Engin Zeydan, Giupponi, L. and Dini, P., Detecting Mobile Traffic Anomalies Through Physical Control Channel Fingerprinting: A Deep Semi-Supervised Approach. 7, doi:https://doi.org/10.1109/access.2019.2947742, pp.152187–152201, 2019.

Biographies

Remalia Mayeni is a Masters Degree student in the Faculty of Applied Science Department of Informatics at the National University of Science and Technology in Zimbabwe. She holds a Bachelor of Technology Honors degree in Computer Science from the Harare Institute of Technology in Zimbabwe. She is presently employed as a Network Engineer within the Core Network division. Her professional background includes expertise in computer science focusing on data networks, security, and cloud computing.

Belinda Mutunhu Ndlovu is a Ph.D. in Information Systems student at UNISA. She holds an MSc in Information Systems and a BSc in Computer Science. She is a seasoned software developer and academic. She has published several papers in the fields of Data Analytics, Health Informatics, ICT4D, and 4IR.

Sibusisiwe Dube is a Professor at the National University of Science and Technology in Zimbabwe. She holds a BSc Hons in Information Systems from the Midlands State University in Zimbabwe, an MSc in Computer Science from the National University of Science and Technology in Zimbabwe and a PhD in Information Systems from the University of Cape Town in South Africa. She has taught courses in Computer Science, Informatics, Information Systems and Data Science at Midlands State University and the National University of Science and Technology in Zimbabwe. Sibusisiwe is an experienced researcher and has published articles in peer-reviewed journals and conference proceedings. She is also an active researcher for articles published In journals such as Global Research & Development Services Publishing, The African Journal of Information Systems, Journal of Information & Management, African Conference on Information Systems and Technology (ACIST), ICT for Africa, Journal of Information Systems Research (INFORMS), International Journal of Research and Innovation in Social Science (IJRISS).

Fungai Jacqueline Kiwa holds a Doctor of Philosophy Degree in Cultural Heritage and Information Technology from Chinhoyi University of Technology, complemented by a Post Graduate Diploma in Higher Education. Additionally, she holds a Master of Science degree in Information Systems, a Bachelor of Technology (Honors) degree in Computing and Information Technology, and an Advanced Diploma in Computing and Information Technology. With a wealth of academic achievements, Dr. Kiwa has authored 17 publications, encompassing articles, conferences, thesis, and dissertations. Currently, she is a candidate for a Master of Mechatronics and AI at the University of Zimbabwe. Her expertise extends to Artificial Intelligence, creative IoT framework designing, and intensive programming skills, particularly in Java, Python, and C++.

Martin Muduva is currently pursuing his PhD at the Chinhoyi University of Technology, Martin is also concurrently engaged in Master's programs in Innovation and Entrepreneurship at the Bindura University of Science Education and Mechatronics and Artificial Intelligence at the University of Zimbabwe. He holds a Master's degree in Leadership and Corporate Governance from Bindura University of Science Education, along with expertise in Big Data Analytics and Information Security. Martin's commitment to education and professional development is evident through his Certificate in Higher and Tertiary Education. His multidisciplinary background and dedication make him a valuable asset in technology, innovation and corporate governance.