FRAUDULENT BANKING TRANSACTIONS DETECTION USING MACHINE LEARNING

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: The vulnerability of banking systems has exposed customers and banks to fraudulent activities, leading to substantial financial losses and reputational damage. Financial fraud in banks is estimated to result in significant annual losses. Early detection is crucial for mitigating fraud, developing counter strategies, and recovering losses. This paper proposes a machine learning-based approach to enhance fraud detection. We analyzed various intelligent algorithms trained on a public dataset to identify correlations between certain factors and fraudulent activities. To address the high class imbalance in the dataset, resampling techniques were employed, and the data was analyzed using the proposed algorithm to achieve better accuracy.

KEYWORDS: Machine learning algorithms, Correlation, Demography, Computational modeling, Finance, Banking Forestry.

1. INTRODUCTION

The banks of the future are very different in terms of their functionalities, compared to them what they are today. These changes are due to the changes in infrastructures, services, people, and skill sets. This transformation is only due to the implementation of financial technologies in banking. Most banks are capable to adopt innovative technologies to deliver financial services and it changes the banking role as we want. New technologies such as blockchain, AI, big data, digital payment processing, peer-to-peer lending, crowd funding, and robot advisors play a vital role in delivering banking services. What is the need for these technological revolutions in banking? As there is a technological evolution, the banking industry is at the forefront of adopting them in their activities to deliver better customer services, but many times the financial crises have adversely affected these new ventures in the banking industry, as a result, innovation was a very distant priority.

At the same time, many new technologies are found as gamechanger for transforming the conventional banking system into

customer-friendly banks. Still, a gap was created between what the bank was offering to its customer and their experience and convenience perspective. Figure (1) represents the different banking activities supported by FinTech companies to improve customer experience by implementing AI technology. This gap was a research topic for many researchers. The traditional banking system is also varied about this technological growth with the expectation and requirements of touch points with the customers with trust and confidence in these technologies. To augment this and provide better technological support there are hundreds of new FinTech companies offering products and services to the banks; p-2-p lending, provides consumer alternatives to loans that were already available in the banks, and robo advisory platform offers to the customers a set of user-friendly solutions.

These services are highly visible and cost-effective. They are very convenient to the consumers with a GUI interface and leave the back-end processing as in conventional banks, such as post-dated settlement, consolidation, and regular reporting. This changes the future banking model by keeping the traditional banking operation at the backend becoming a commoditized utility provider. A technological front and the front end control the customer experience. This technological innovation in banking is also connected to several other positive developments in the related industrial segment. Fig.1. AI Technology to improve customer experience in Banking Activities AI-powered chatbots that mimic human conversation and messaging apps are replacing the activities of the backend services in call centers. Biometric data and iris scanning are used as an alternative to passwords and tokens used for transactions.

2. LITERATURE SURVEY

Statistical methods can be used for fraud detection. Here the statistic distribution of the dataset is analysed for anomalous behaviour of the fraudulent by using Linear Discriminate Analysis and Logistic regression [1]. The author used a variety of data mining techniques in real-time fraud detection using historical data [1]. The research work [2] describes the methods to detect fraud by using KNN algorithm and outlier finding mechanism. The model helps in the detection of malicious behavior of the fraudulent. The authors in [3] used an ensemble technique including the Random Forest model to analyze the normal transactions and compare the performance of the fraudulent transaction detection

method by neural networks. Fraud detection in [4] presented the method for credit card transactions and analyzed the data using Wale-algorithm optimized backpropagation. The authors in [4][6] have analyzed already classified results for detecting credit card fraud using an imbalanced dataset. K means clustering is used for sampling groups of fraudulent transaction samples. Authors also used genetic algorithms for group fraudulent transactions.

3. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

The existing system for "Fraud Detection in Banking Transactions Using Machine Learning" incorporates a comprehensive approach to mitigating financial fraud within the banking sector. Initially, historical transaction data is collected, encompassing a diverse range of transactions, and undergoes rigorous pre processing to handle missing data, address imbalances, and normalize features. The exploratory data analysis phase provides critical insights into patterns and correlations. Following this, relevant features are carefully selected to contribute to the fraud detection process. The model development phase employs machine learning algorithms, with a focus on continuous optimization through hyperparameter tuning. The AI-based model is implemented within the banking system for real-time or batch processing of transactions. Evaluation metrics, including accuracy, precision, recall, and AUC-ROC, are employed to assess the model's performance. Continuous monitoring mechanisms and feedback loops are established for adaptive improvements, ensuring the model remains effective against evolving fraudulent activities. The entire process is thoroughly documented, providing insights into data sources, preprocessing steps, model development, and evaluation metrics. Furthermore, security measures are integrated to safeguard both the model and the sensitive financial data it processes, encompassing encryption, access controls, and other relevant security best practices.

LIMITATIONS OF EXISTING SYSTEM

Evolution of Fraud Patterns: Fraudulent activities evolve over time, and the model may not adapt quickly enough to new types of fraud. Regular updates and continuous monitoring are crucial to ensuring the model's effectiveness against emerging threats.

Over fitting: Over fitting occurs when a model performs well on the training data but fails to generalize to new, unseen data. This

can lead to a high level of accuracy on the training set but reduced performance on real-world scenarios.

3.2 PROPOSED SYSTEM

The proposed system for "Fraud Detection in Banking Transactions Using Machine Learning" aims to overcome the limitations of the existing system by introducing innovative strategies and technologies. To address imbalanced data issues, the proposed system employs advanced resampling techniques to mitigate biases and enhance the model's ability to detect instances of fraud across various classes. A key focus lies in the continuous evolution of the fraud detection model to adapt to emerging patterns through regular updates facilitated by a dynamic learning mechanism. To mitigate overfitting, the proposed system integrates sophisticated regularization techniques and explores ensemble methods to improve the model's generalization to unseen data. Interpretability is enhanced through the incorporation of explainable AI techniques, Regulatory compliance is integrated into the core of the proposed system, ensuring adherence to legal and ethical standards. User acceptance is fostered through comprehensive training and communication strategies to instill confidence in the reliability and effectiveness of the machine learning-based fraud detection system. Through these advancements, the proposed system aims to not only enhance the accuracy and efficiency of fraud detection but also ensure adaptability, transparency, and compliance in the ever-evolving landscape of banking transactions.

4. SYSTEM ARCHITECTURE

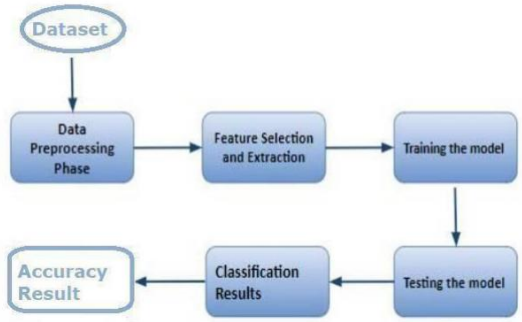

Fig 4: System Architecture

5. METHODOLOGY

Most banks adopt traditional rule-based methods of fraud analysis. Today due to the availability of advanced technologies the number of fraudsters is increasing, which is also an increased threat level to the banking industry. Fraud patterns are changing due to inconsistency in the banking systems. Fraud detection is possible with a valuable dataset and a high-performance machine learning algorithm. The data are gathered from a public dataset and categorized, based on these we can classify the users as benign or fraudulent gives the details about the fraud

detection and prevention market size in 2016 – 2022, worldwide. Many statistical and machine learning models are used to analyze the fraudulent and non-fraudulent in each dataset. In this paper, we analyze popular statistical and machine-learning methods for the detection of fraudulent transactions. The most popular among these is Benford's law for modelling and the other machine learning modules for classification and binary decision trees [12]. These models help to determine benign and fraudulent transactions.

Under machine learning determining whether the transaction is fraudulent or benign is considered a classification problem. Different machine learning algorithms play a crucial role in fraud detection[21]. This includes Logistic regression, k-nearest neighbour algorithms, Random Forest (RF) Classifier, Support Vector Machine (SVM), and Naïve Bayes classifier.

6 . MODULES

i)Data Preprocessing Module: This module handles the collection, cleaning, and preparation of the dataset. It includes tasks such as handling missing data, addressing imbalances, normalizing numerical features, encoding categorical variables, and splitting the data into training and testing sets. The goal is to ensure that the data is in a suitable format for training the machine learning models.

ii)Feature Engineering and Selection Module: This module focuses on selecting and transforming relevant features from the dataset to improve the model's performance. Techniques such as correlation analysis, feature importance scoring, and dimensionality reduction are employed to identify and extract the most informative features for fraud detection.

iii)Machine Learning Model Development Module: The core of the system, this module involves choosing, training, and fine-tuning machine learning algorithms for fraud detection. Decision trees, random forests, support vector machines, or neural networks can be explored. Hyper parameter tuning is conducted to optimize the model's performance, and the trained model is integrated into the system.

7. RESULT

Fig 7.1: Bank Transaction Prediction by using Machine Learning

We can see from the above data that only *two* type of transactions are classified as fraud so we will drop the remaining types to generalize the data and we will only keep Cash_out and Transfer type.

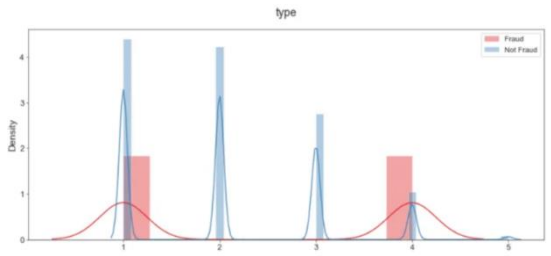

Fig 7.2: Transactions are classified as Fraud or not Fraud

The Type feature in our data is categorical so we will map it to convert it to numerical data 6,3544,407 transactions were Not Fraud transactions with 2762196 Not Fraud transactions after considering only two types which are relevant with only 0.3% Fraud transactions. This shows us that we have a very imbalanced data.

8. CONCLUSION&FUTURE SCOPE

The research proposes the integration of Random Forest, K-Nearest Neighbours (KNN), and Logistic Regression algorithms for fraud detection in banking applications. Utilizing a publicly available dataset from UCI, the analysis reveals a significant imbalance biased towards the majority of samples, addressed by the Synthetic Minority Over-sampling Technique (SMOTE). Implementation challenges with KNN and Random Forest method. The model demonstrates a commendable 97.74% performance rate. Notably, the analysis highlights a higher likelihood of fraudulent activity among individuals aged 19-25. This suggests the importance of considering demographic factors in fraud detection. The proposed ensemble of Random Forest, Logistic Regression and KNN offers an effective solution to address dataset imbalances and enhance fraud detection accuracy in banking applications Future studies and work, we propose using other hybrid models as well as working specifically in the field of Cat Boost by changing more hyper parameters, especially the hyper parameter number of trees. Also, due to hardware limitations in this study, the use of stronger and better hardware may bring better results that can ultimately be compared with the results of this study.

REFERENCES

[1] R. Rambola, P. Varshney and P. Vishwakarma, “Data Mining Techniques for Fraud Detection in Banking Sector,” 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-5, doi: 10.1109/CCAA.2018.8777535.

[2] N. Malini and M. Pushpa, "Analysis on credit card fraud identification techniques based on KNN and outlier detection," 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, 2017, pp. 255-258, doi: 10.1109/AEEICB.2017.7972424.

[3] Ishan Sohony, Rameshwar Pratap, and Ullas Nambiar. 2018. Ensemble learning for credit card fraud detection. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (CoDS-COMAD '18). Association for Computing Machinery, New York, NY, USA, 289–294. DOI:https://doi.org/10.1145/3152494.3156815

[4] C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai, and S. Pan, "Credit Card Fraud Detection Based on Whale Algorithm Optimized BP Neural Network," 2018 13th International Conference on Computer Science Education (ICCSE), Colombo, 2018, pp. 1-4, doi: 10.1109/ICCSE.2018.8468855

[5] I. Benchaji, S. Douzi and B. ElOuahidi, "Using Genetic Algorithm to Improve Classification of Imbalanced Datasets for Credit Card Fraud Detection," 2018 2nd Cyber Security in Networking Conference (CSNet), Paris, 2018, pp. 1-5, doi: 10.1109/CSNET.2018.8602972.

[6] John O. Awoyemi, Adebayo Olusola Adetunmbi, and Samuel Adebayo Oluwadare. Credit card fraud detection using machine learning techniques: A comparative analysis. 2017 International Conference on Computing Networking and Informatics (ICCNI), pages 1–9, 2017.

[7] Fabrizio Carcillo, Andrea Dal Pozzolo, Yann-A¨el Le Borgne, Olivier Caelen, Yannis Mazzer, and Gianluca Bontempi. Scarff: a scalable framework for streaming credit card fraud detection with spark. Information Fusion, 41:182–194, 2018.

[8] Galina Baader and Helmut Krcmar. Reducing false positives in fraud detection: Combining the red flag approach with process mining. International Journal of Accounting Information Systems, 2018.

[9] Ravisankar P, Ravi V, Raghava Rao G, and Bose, Detection of financial statement fraud and feature selection using data mining techniques, Elsevier, Decision Support Systems Volume 50, Issue 2, p491-500 (2011) SVM

[10] K. Seeja, and M. Zareapoor, "FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining," The Scientific World Journal,2014, pp. 1-10.