International Journal for Modern Trends in Science and Technology
Volume 10, Issue 04, pages 315-320.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol10issue04.html
DOI: https://doi.org/10.46501/IJMTST1004047

A Machine Learning Approach to Combatting Fraudulent Activities in Banking Data

T.N.V Durga[1] , Kalaga Rama Latha[2], Mallidi Jayasri[2], Setti Venkata Kalyani[2], Nakka Khaga Akhil[2], Mummana Manjunadh[2]

[1]Assistant Professor, Department of Computer Science Engineering, Pragati Engineering College, Surampalem , Andhra Pradesh, India.
[2]Department of Computer Science Engineering, Pragati Engineering College, Surampalem , Andhra Pradesh, India.

To Cite this Article
T.N.V Durga, Kalaga Rama Latha, Mallidi Jayasri, Setti Venkata Kalyani, Nakka Khaga Akhil, Mummana Manjunadh, A Machine Learning Approach to Combatting Fraudulent Activities in Banking Data, International Journal for Modern Trends in Science and Technology, 2024, 10(04), pages. 315-320. https://doi.org/10.46501/IJMTST1004047

Article Info
Received: 06 April 2024; Accepted: 18 April 2024; Published: 26 April 2024.

ABSTRACT

Banking system vulnerabilities have made us vulnerable to fraudulent activities that seriously harm the bank's brand and financial standing in addition to harming clients. An estimated large sum of money is lost financially each year as a result of financial fraud in banks. Early discovery aids in the mitigation of the fraud by allowing for the development of a countermeasure and the recovery of such losses. This research proposes a machine learning-based method to effectively aid in fraud detection. In order to combat counterfeits and minimize damage, an AI oriented system will expedite the check verification process. In order to determine the relationship between specific characteristics and fraudulence, we examined a number of clever algorithms that were trained on the publicly available data in this article. The dataset used in this study is resampled to reduce the high class of instability in it, and the suggested technique is used to analyse the data for more accuracy.

Keywords: Seismic Image, Image Segmentation, DCNN, Auto-Encoder

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, crowdfunding, 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. The paper is structured with the following sections. In section II we described the literature review with the related work completed by other researchers and in section III technological impact on banking and the digital revolution in India. Section IV describes the role of AI in risk management and governance and section V, the fraud analysis using machine learning algorithms followed by a conclusion.

2. LITERATURE SURVEY

Researchers have suggested a number of techniques to stop fraudulent transactions and identify credit card fraud. Below is a summary of recent, relevant works that are at the cutting edge. A novel model known as the AIS-based fraud detection model (AFDM) is studied by Halvaiee and Akbari. They increase the accuracy of fraud detection by utilizing the Immune System Influenced Algorithm (AIRS). According to their paper's results, the suggested AFDM can outperform simple algorithms in terms of accuracy, cost savings, and system reaction time, with improvements of up to 25%, 85%, and 40%, respectively. Bahnsen et al. used the von Mises distribution to build a transaction aggregation technique and a new set of characteristics based on the periodic behaviour analysis of the exchange time. Furthermore, they present a novel cost-based criterion for assessing the models used in credit card fraud detection, and they analyse the impact of various feature sets on outcomes using an actual credit card dataset. To be more exact, they expand on the transaction aggregation approach by developing new offers that are determined by examining the periodic patterns in transaction behaviour. The use of machine learning algorithms to identify credit card fraud is studied by Randhawa et al. Initially, they assess the available datasets using Naïve Bayes, conventional models for support vector machines, neural networks, logistic regression, linear regression (LR), and stochastic forest and decision trees. Additionally, they suggest a hybrid approach that combines majority voting with AdaBoost. Furthermore, they introduce noise into the data samples in order to assess their resilience. Using publicly accessible statistics, they conduct tests and demonstrate the efficacy of majority voting in identifying instances of credit card fraud. Porwal and Mukund present a robust method for finding outliers in a big dataset by using clustering techniques. to shifting trends. Their suggested method is predicated on the ideas that users' excellent conduct is stable as time passes and that the information sets that demonstrate positive behaviour have a similar spatial signature across various groups. They demonstrate how spotting modifications to this data may be used to identify fraudulent activity. They demonstrate that when used as an assessment criterion, the area under the curve of precision recall performs better than ROC. A group learning approach based on training set partitioning and clustering is proposed by the authors in. Their suggested approach aims to address the dataset's extreme imbalance in addition to preserving the validity of the sample characteristics. Their suggested framework's primary feature is its ability to train each base estimator in

simultaneously, which increases the framework's efficacy.

3. SYSTEM ANALYSIS

A. EXISTING SYSTEM

The present "Fraud Detection in Banking Transactions Using Machine Learning" system incorporates a comprehensive plan for eliminating financial fraud in the banking sector. The initial stage is to collect historical transaction information, which covers a wide range of transactions. The data is then thoroughly preprocessed to correct for imbalances, handle missing data, and standardize features. Exploratory data analysis provides valuable insights on patterns and correlations. Following that, factors related to the identification of fraud technique are carefully selected. During the model generation phase, machine learning approaches are used, with a focus on hyperparameter modification for continuous optimization. The banking system utilizes an AI-based model to process payments in batches or in real time. The model's performance is evaluated using AUC-ROC, precision, recall, precision, and other metrics. Mechanisms for continuous monitoring and feedback loops are established for adaptive upgrades, ensuring that the model remains effective against changing fraudulent behaviors. The entire technique is extensively documented, including details on the model's design, preprocessing phases, data sources, and evaluation measures. Furthermore, access controls, encryption, and other relevant security standards are built into the security mechanisms to secure the model and the highly confidential financial data it handles.

DISADVANTAGES OF THE EXISTING SYSTEM

Imbalanced Data Issues:

If the dataset used for training is highly imbalanced, where instances of fraud are significantly outnumbered by non-fraudulent transactions, the model may have a bias toward the majority class, potentially leading to lower sensitivity in fraud detection.

Evolution of Fraud Patterns:

Over time, fraudulent actions change, and the model might not be able to keep up with these developments in fraud. Constant monitoring and regular updates are essential to guaranteeing the model's efficacy against new threats.

Overfitting:

When a model works well on training data but is unable to generalize to fresh, untried data, this is known as overfitting. This may result in decreased performance in real-world circumstances yet a high degree of accuracy on the training set.

False Positives and False Negatives:

The model could provide false negatives (missing real cases of fraud) or false positives (erroneously labeling non-fraudulent transactions as fraudulent). It is difficult to balance these faults, and constant modifications are needed to reduce both kinds of errors.

Interpretability:

Complex machine learning models, such as deep neural networks, might lack interpretability, making it difficult to understand how and why a particular decision was made. Interpretability is crucial in the context of financial transactions, where explanations for flagged activities are essential.

Data Quality and Variability:

Incomplete or poor-quality data can negatively impact the model's performance. Additionally, variations in data quality over time or across different sources may introduce challenges in maintaining a consistently high level of accuracy.

Computational Resources:

Resource-intensive models may require significant computational power, leading to increased processing times and costs. This can be a limitation in scenarios where real-time processing of transactions is essential.

Adversarial Attacks:

Sophisticated attackers may attempt to manipulate the model by providing adversarial input designed to mislead the system. Ensuring resilience against such attacks is a continuous challenge in the field of fraud detection.

Regulatory Compliance:

Compliance with regulatory requirements and data protection laws, such as GDPR, may pose

challenges. Ensuring that the model adheres to legal and ethical standards is essential in the banking sector.

User Acceptance:

Users within the banking system might be skeptical or resistant to fully trusting machine learning models for critical tasks. Ensuring user acceptance and understanding is crucial for successful implementation.

B. 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, ensuring that stakeholders can comprehend and trust the decision-making process of the model. Additionally, the proposed system places a strong emphasis on data quality and variability, implementing robust data validation and cleansing protocols. To address computational resource constraints, optimization strategies are explored to enhance the efficiency of processing, ensuring timely and cost-effective fraud detection. In order to strengthen resilience against adversarial assaults, the system also includes robust security procedures that guard against input manipulation. The suggested system incorporates regulatory compliance into its core to guarantee that legal and ethical requirements are followed. Extensive training and communication tactics are used to build user acceptability and instil trust in the efficacy and dependability of the machine learning-based identification of fraud system. By utilizing these developments, the suggested method hopes to guarantee flexibility, openness, and compatibility in the constantly changing world of financial transactions, in addition to improving the precision and effectiveness of fraud detection.

ADVANTAGES OF THE PROPOSED SYSTEM

The proposed system for "Fraud Detection in Banking Transactions Using Machine Learning" offers several advantages over the existing system:

Improved Detection Accuracy:

Leveraging advanced machine learning algorithms and dynamic learning mechanisms, the proposed system enhances the accuracy of fraud detection. This improvement is crucial in minimizing both false positives and false negatives, ensuring a more precise identification of fraudulent transactions.

Adaptability to Emerging Threats:

The proposed system's emphasis on continuous evolution through regular updates enables it to adapt swiftly to emerging fraud patterns. This adaptability is essential in the ever-changing landscape of financial fraud, allowing the system to stay ahead of new and sophisticated techniques employed by malicious actors.

Enhanced Interpretability:

By incorporating explainable AI techniques, the proposed system provides stakeholders with a clearer understanding of the decision-making process. Enhanced interpretability fosters trust among users, auditors, and regulatory bodies, addressing concerns related to the transparency of the fraud detection model.

Optimized Resource Utilization:

Optimization strategies integrated into the proposed system enhance computational efficiency, enabling timely processing of transactions without compromising accuracy. This advantage is particularly valuable in real-time processing scenarios, where swift identification of fraudulent activities is paramount.

Comprehensive Security Measures:

The proposed system incorporates advanced security measures to fortify resilience against adversarial attacks and protect against manipulative inputs. By addressing potential

vulnerabilities, the system ensures the integrity and confidentiality of sensitive financial data, maintaining a robust defense against malicious activities.

4. SYSTEM DESIGN

SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture.

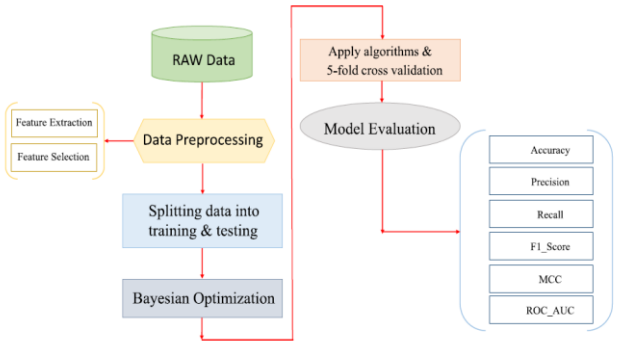

Fig 1. Methodology followed for proposed model

5. SYSTEM IMPLEMENTATION

MODULES

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.

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.

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. Hyperparameter tuning is conducted to optimize the model's performance, and the trained model is integrated into the system.

Continuous Learning and Update Module:

This module ensures the adaptability of the system to evolving fraud patterns. It involves mechanisms for continuous learning, where the model is regularly updated with new data to stay relevant and effective. Feedback loops are established to incorporate insights from false positives and false negatives, facilitating ongoing improvements in fraud detection accuracy.

Security and Compliance Module:

This module addresses the security and compliance aspects of the system. It includes measures to safeguard the model and data from adversarial attacks, encryption of sensitive information, access controls, and compliance with regulatory standards such as GDPR. Security protocols are integrated to protect the integrity and confidentiality of financial data processed by the system.

6. RESULTS AND DISCUSSION

We use the stratified 5-fold cross validation method and the boosting algorithms with the Bayesian optimization method to evaluate the performance of the proposed framework. We extract the hyperparameters and evaluate each algorithm individually before using the majority voting method. We examine the algorithms in triple and double precision. The comparison results are presented in Table.

Model	Accuracy	AUC	Recall	Precision	F1-score	MCC
Log_Reg	0.97477	0.9578	0.8730	0.0617	0.1143	0.2248
LGBM	0.99919	0.9472	0.7990	0.7534	0.7699	0.7727
XGB	0.99923	0.9517	0.7949	0.7862	0.7830	0.7864
CatBoost	0.99880	0.9390	0.8096	0.6431	0.7066	0.7158
Vot_Lg, Xg, Ca	0.99924	0.9501	0.8033	0.7720	0.7825	0.7847
Vot_Lg, Xg	0.99927	0.9522	0.8012	0.7901	0.7901	0.7925
Vot_g, Ca	0.99923	0.9492	0.8097	0.7681	0.7823	0.7852
Vot_Lg, Ca	0.99912	0.9459	0.8075	0.7260	0.7581	0.7620

Table 1. **Performance evaluation of algorithms.**

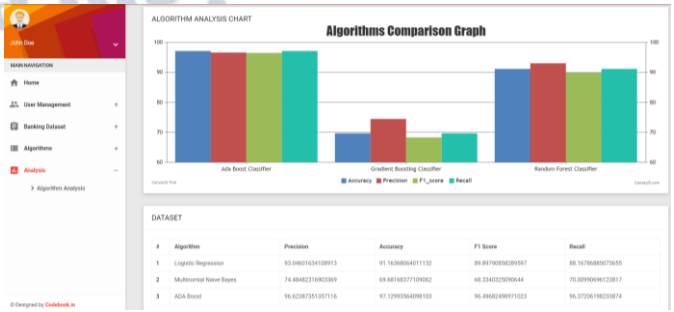

Fig 2. Algorithms Comparisons

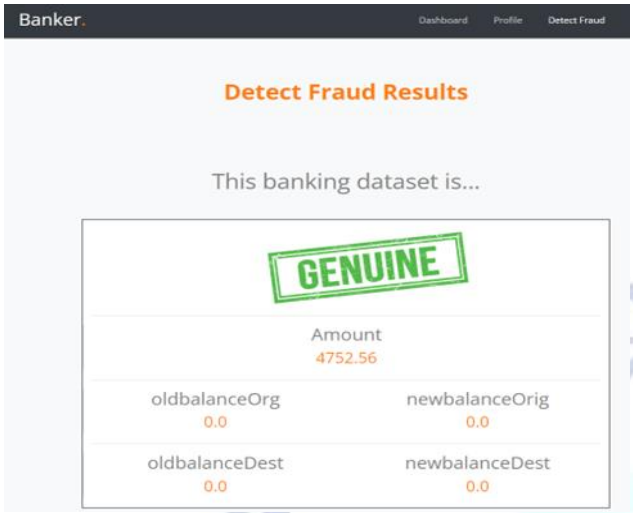

***Fig 3.* Detect Fraud Transactions-1.**

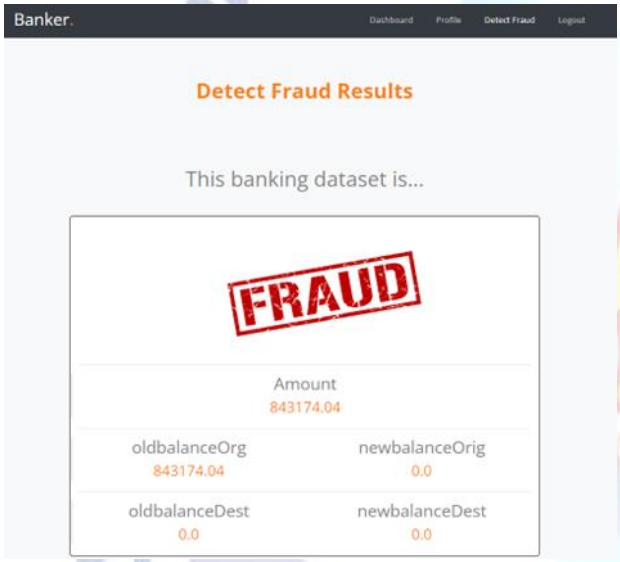

fig 4. Detect Fraud Transactions-2

7. CONCLUSION AD FUTURE WORK

Use of ML algorithms proposed in this research to detect fraud in banking applications. The publicly available dataset from UCI is analysed. The high level of imbalance in the dataset provided is highly biased toward the majority of samples. This problem is tackled by the synthetic minority over-sampling technique (SMOTE). Implementation issues of this by KNN and Random Forest algorithms are handled by XGBoost as the boosting methods. The performance achieved using the model was 97.74%. In the analysis of the dataset, we found that people in the age group of 19-25 years are more likely to be fraudulent than other customers' demography.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

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. KNN, SVM

[11] C. Tyagi, P. Parwekar, P. Singh, and K. Natla, "Analysis of Credit Card Fraud Detection Techniques," Solid State Technology, vol. 63, no. 6, 2020, pp. 18057-18069.