**FRAUD DETECTION IN BANKING TRANSACTIONS USING MACHINE LEARNING**

**ABSTRACT**

Vulnerability in banking systems has exposed us to fraudulent acts, which cause severe damage to both customers and the bank in terms of loss of money and reputation. Financial fraud in banks is estimated to result in a significant amount of financial loss annually. Early detection of this helps to mitigate the fraud, by developing a counter strategy and recovering from such losses. A machine learning-based approach is proposed in this paper to contribute to fraud detection successfully. The artificial intelligence (AI) based model will speed up the check verification to counteract the counterfeits and lower the damage. In this paper, we analyzed numerous intelligent algorithms trained on a public dataset to find the correlation of certain factors with fraudulence. The dataset utilized for this research is resampled to minimize the high class of imbalance in it and analyzed the data using the proposed algorithm for better accuracy.

**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. 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.

The other technologies enabled with FinTech are IoT, wearable technologies and ben-in-banking are common things in day-to-day banking operations, most of them with gamification service to the end users. To service their customers, banks today need to change their mode of service. By adopting advanced technologies, they may succeed in the evolution of the banking industry and embed them in their operations as their culture and innovation across the organization. This has consequences in many folds. The City bank analysis represents that, for the next 10 years 30% of the bank jobs will disappear. Some research estimated this job loss will be more than 50%. This has a consequence for many financial institutions across the globe. It is not just a job loss, but also connected with several economic aspects around it, like accounting firms, law firms, hotels, and services businesses. The profiles of the new jobs created in the FinTech industry are very small in number, but they are entirely different with very different skill sets.

Those are very essential for today’s banking systems. Considering all the above-said challenges in banking one of the common challenges faced by most of the banks today are fraudulent transactions and malicious behavior of the users. The fraudulent transaction is a rising challenge for the banking industry with an estimated financial loss and damage to the reputation. This survey report presents that about 56% of the fraud in banking is only reported. It clearly indicates that a better fraud detection 978-1-6654-9260-7/23/$31.00 c 2023 IEEE 221 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE) | 978-1-6654-9260-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IITCEE57236.2023.10091067 Authorized licensed use limited to: Zhejiang University. Downloaded on November 06,2023 at 16:28:16 UTC from IEEE Xplore. Restrictions apply. model is of paramount requirement in most banks either small or big in size. This research work aims to design an intelligent system with a machine learning model, that will be predictive and adaptive to detect the fraudulent activities of the customer in banking transactions.

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.

**SYSTEM ANALYSIS**

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 preprocessing 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 hyper parameter 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**

**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:**

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.

**Overfitting:**

Overfitting 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.

**False Positives and False Negatives:**

The model may produce false positives (incorrectly flagging non-fraudulent transactions as fraudulent) or false negatives (missing actual instances of fraud). Balancing these errors is challenging, and it requires ongoing adjustments to minimize both types of mistakes.

**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.

**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. The system also incorporates mechanisms to fortify resilience against adversarial attacks, leveraging advanced security measures to protect against manipulative inputs. 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.

**ADVANTAGES OF 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.

**SYSTEM ARCHITECTURE**



**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.

**HARDWARE REQUIREMENTS**

|  |  |  |
| --- | --- | --- |
| MINIMUM (Required for Execution) | | MY SYSTEM (Development) |
| System | Pentium IV 2.2 GHz | i3 Processor 5th Gen |
| Hard Disk | 20 Gb | 500 Gb |
| Ram | 1 Gb | 4 Gb |

**SOFTWARE REQUIREMENTS**

|  |  |
| --- | --- |
| Operating System | Windows 10/11 |
| Development Software | Python 3.10 |
| Programming Language | Python |
| Domain | Image Processing & Cloud Computing |
| Integrated Development Environment (IDE) | Visual Studio Code |
| Front End Technologies | HTML5, CSS3, Java Script |
| Back End Technologies or Framework | Django |
| Database Language | SQL |
| Database (RDBMS) | MySQL |
| Database Software | WAMP or XAMPP Server |
| Web Server or Deployment Server | Django Application Development Server |
| Design/Modelling | Rational Rose |

**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.