International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

A Predictive Model for Detecting Fraudulent Transactions in Financial Systems

S. A. Bagul[1], Vaibhavi Handibag[2], Yash Lawande[3], Prem Mandhare[4], Shrikant Nevse[5]

[1]Professor, Department of Information Technology, JSCOE

[2, 3, 4, 5]Department of Information Technology, JSCOE

Abstract: ***Global financial security is seriously threatened by the rise in fraudulent activities brought on by the development of digital financial systems. Conventional rule-based systems are unable to adjust to changing fraud trends. The Light Gradient Boosting Machine (LightGBM) algorithm is used in this paper's machine learning-based financial fraud detection system, which is coupled with a Streamlit dashboard for real-time prediction and visualization. The system employs AutoML for hyperparameter optimization and SMOTE for data balancing. The trained LightGBM model accurately and interpretably classifies transactions as either fraudulent or legitimate. The interactive Streamlit interface offers analytical insights and visualizes fraud trends. This model is appropriate for real-world financial fraud detection because it exhibits scalability, transparency, and real-time responsiveness.***

Keywords: Fraud Detection, LightGBM, Machine Learning, Streamlit, SMOTE, AutoML, Digital Security

1. Introduction

The increasing use of digital and online payment methods has coincided with a rise in fraudulent activity, including account takeovers and credit card abuse. It is estimated that financial fraud costs the world more than $35 billion every year. Conventional methods of detecting fraud rely on rule-based systems, which are unable to identify new and flexible fraudulent patterns.

This study suggests an AI-based predictive model that uses LightGBM in conjunction with Streamlit for analytics and visualization in order to get around these problems. The model uses AutoML optimization, SMOTE to handle imbalanced data, and SHAP analysis to generate results that can be understood. The goal of this research is to develop a fraud detection framework that is transparent, high-performing, and easy to use for practical applications. Financial systems have become more convenient as a result of their quick digitization, but they are also more vulnerable to fraud risks like identity theft, unauthorized withdrawals, and credit card abuse. The 2024 Nilson Report states that losses from card fraud worldwide topped $35 billion Real-time detection of these fraudulent activities is essential to guaranteeing safe transactions and client confidence. Because they are static, traditional rule-based systems are unable to adjust to changing fraud patterns. This project suggests a LightGBM-based fraud detection system that can handle unbalanced data, learn intricate transaction patterns, and provide real-time predictions in order to get around these restrictions.

2. Literature Survey

1) Title of Paper: Credit Card Fraud Detection: A Novel Learning Approach and a Realistic Modeling
Authors: G. Bontempi, O. Caelen, and R. Dal Pozzolo
Summary:
In this paper, time-based learning and adaptive evaluation techniques are used to present a realistic modeling framework for credit card fraud detection.

- Tackles significant issues in fraud detection, including concept drift, class imbalance, and verification latency.
- Suggests using a time-based data split to model evaluation in order to replicate real-world conditions.
- Makes use of a dataset that spans three years and includes over 75 million transactions.
- Shows how fraud detection performance is enhanced by adaptive learning and appropriate evaluation metrics.

2) Title of Paper: Attention-Based Gated Network for Credit Card Fraud Detection with Time Awareness
Authors: X. Zhang, C., and Y. Li are the authors. Lin

Summary:
- A Time-Aware Attention-Based Gated Network (TAGN) is presented in this paper to identify behavioral and temporal patterns in user transactions.
- Uses time-aware gates to learn both short-term and long-term user transaction patterns.
- Focuses on important behavioral changes by using attentional mechanisms.
- Enhances the accuracy of fraud recognition by performing well on unbalanced sequential datasets.
- Surpasses RNN and CNN-based techniques in terms of recall and F1-score.
- Demonstrates strong generalization ability across multiple real world credit card fraud datasets.

3) Title of Paper: Enhanced LightGBM for Highly Unbalanced Data and Use in Credit Card Fraud Identification
Authors: Z. Zhang, F. Wang, and Y. Chen

Summary:
In this work, the Light Gradient Boosting Machine (LightGBM) framework for detecting credit card fraud in the presence of severe data imbalance is improved.
- Presents two variations: Oversampling Cost-Harmonization LightGBM (OS-CHL) and Class-Balancing Cost-Harmonization LightGBM (CB-CHL).
- Enhances model performance by combining cost-sensitive

learning, class weighting, and oversampling.
- Shows a steady increase in recall and F2-score across several datasets.
- Confirms that LightGBM is a scalable and successful method for detecting fraud on a large scale.

4) Title of Paper: Value-at-Risk and Machine Learning for Financial Fraud Detection in Skewed Data
Authors: M. R. Oloyede, A. O. Adepoju, and A. O. Falade

Summary:
In order to manage skewed fraud datasets, this study investigates Value-at-Risk (VaR) as a financial risk metric.
- Gives transactions risk-based weights determined by the possibility of monetary losses.
- Uuses VaR thresholds to model fraud cases as worst-case situations.
- Achieves a 0.95 true positive rate by implementing a K-Nearest Neighbor (KNN) classifier.
- Presents a novel Detection Rate (DT) metric for assessing the performance of fraud risk
- Presents the Detection Rate (DT) metric, which uses risk-weighted accuracy instead of conventional classification metrics to assess fraud models.
- Shows how combining financial risk assessment and predictive analytics with VaR and machine learning improves fraud detection.

5) Title of Paper: Explainable AI for Financial Fraud Detection Using XGBoost and SHAP
Authors: L. Verma, R. Sharma, and M. Patel

Summary:
- Interprets model predictions by integrating Explainable AI (XAI) techniques with SHAP values.
- Effectively classifies and visualizes important fraud indicators using XGBoost.
- Contributes explainable features to fraud detection models, increasing their transparency and credibility.
- Offers a framework for visual analysis to comprehend the variables affecting the risk of fraud.
- Shows how XAI and machine learning can be combined to enhance interpretability without sacrificing accuracy.

3. Methodology

3.1 Existing System

To find fraudulent transactions in massive financial datasets, the current financial fraud detection systems use a variety of machine learning and deep learning models.

1) **Time-Based Modeling Techniques:** By taking into account the temporal order of transactions, these techniques replicate actual fraud detection situations. Time-based learning techniques were developed by researchers like Dal Pozzolo et al. to address concept drift and verification latency, increasing detection accuracy over time.
2) **Deep Learning with Attention Mechanisms:** Time-Aware Gated Networks and other neural models use attention mechanisms to record user behavior and transaction history in sequential order. Large datasets and a lot of processing power are needed for these models to learn intricate temporal dependencies, which makes real-time deployment more difficult.
3) **Real-Time Translation:** The app employs Google Cloud Translation API to fetch translations instantly. Captured text or object labels are processed and translated into the user's target language, ensuring a smooth and efficient translation experience.
4) **Gradient Boosting-Based Solutions:** To effectively identify infrequent fraudulent cases, frameworks like Improved LightGBM use cost-harmonization and class balancing. Financial analysts find it challenging to interpret model predictions due to their lack of transparency, even though they perform well on datasets that are unbalanced.
5) **Risk-Based and Statistical Techniques:** Value-at-Risk (VaR)-based techniques identify high-risk transactions and estimate possible financial losses. These models are difficult to adjust for real-time fraud detection and rely significantly on historical assumptions, despite being useful for risk quantification. Instead of being incorporated into real-time fraud prevention systems, these solutions are frequently used in post-analysis.

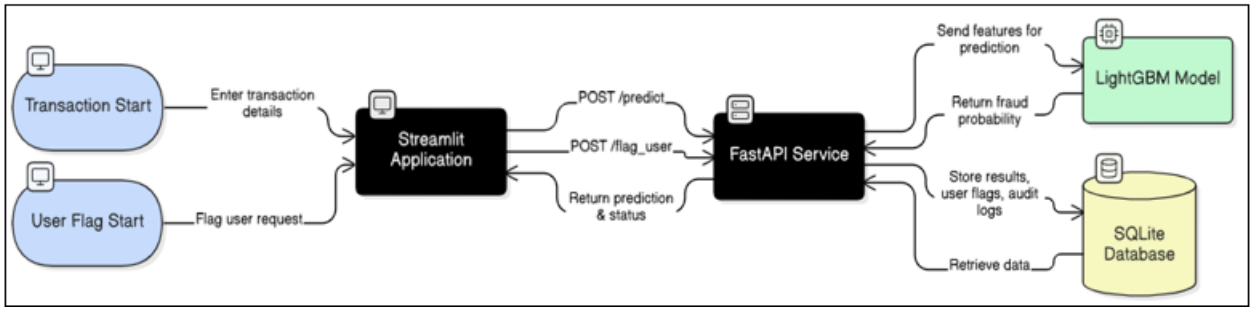

Figure 1: The architecture of system

3.2 Proposed System

With the help of sophisticated machine learning and explainable AI tools integrated into a Streamlit-based platform, this system makes it possible to detect financial fraud in real time and to predict and visualize suspicious transactions.

1) **Streamlit Dashboard User Interface:**
The serves as the primary gateway through which users can upload transaction data or watch live streams.
2) **Data Collection and Preprocessing Module:**
Gathers transaction information, including details like the transaction type, location, amount, and merchant ID.
3) **Module for Data Balancing (SMOTE):**

Creates synthetic samples for minority (fraud) cases using the SMOTE technique to address class imbalance.

4) **The LightGBM Fraud Detection Module:**
Uses the Light Gradient Boosting Machine (LightGBM) algorithm to determine whether a transaction is authentic or fraudulent.
5) **Explainable AI Module (LIME & SHAP):**
Determines which transaction features had the biggest influence on each choice in order to interpret the model's predictions.
6) **Fraud Warning and Document Administration:**
Transactions with a high likelihood of fraud are automatically flagged and stored in a secure database.
Allows for prompt action to stop financial losses by sending alerts for instant review.

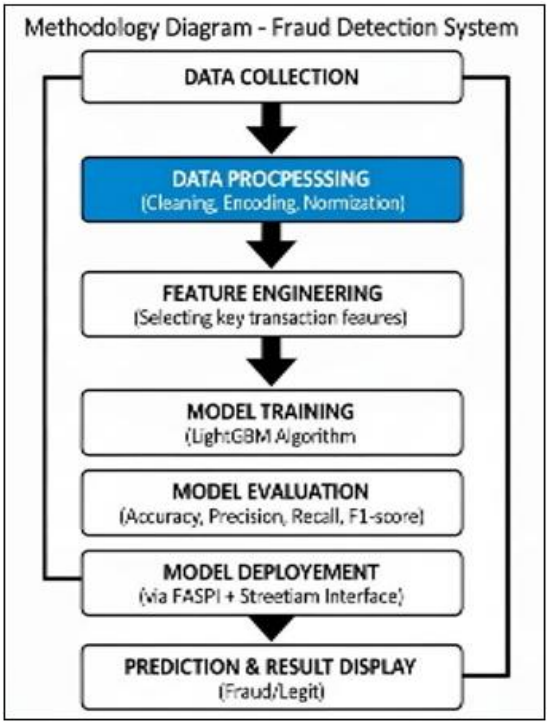

Figure 2: Flowchart

3.3 Detection of Fraud

- Microsoft created the high-performance machine-learning algorithm known as LightGBM (Light Gradient Boosting Machine) to make quick and precise predictions.
- Makes use of gradient-based decision trees, which reduce computation time without sacrificing classification accuracy.
- Uses parameters like class weighting and leaf-wise tree growth to handle unbalanced datasets in an efficient manner.
- A fraud probability score that classifies each transaction as either legitimate or possibly fraudulent is one of the outputs.

Figure 3: Snapshot of Home Page

3.4 Key Technologies

- The suggested system builds an effective and sophisticated fraud detection framework by utilizing LightGBM, SMOTE, and Streamlit.
- Using extensive data, LightGBM, a potent gradient boosting algorithm, learns intricate and obscure fraud patterns to reliably classify financial transactions.
- SMOTE (Synthetic Minority Oversampling Technique) is used to address the problem of data imbalance.
- This combination guarantees a modern financial fraud detection solution that is accurate, scalable, and easy to use.

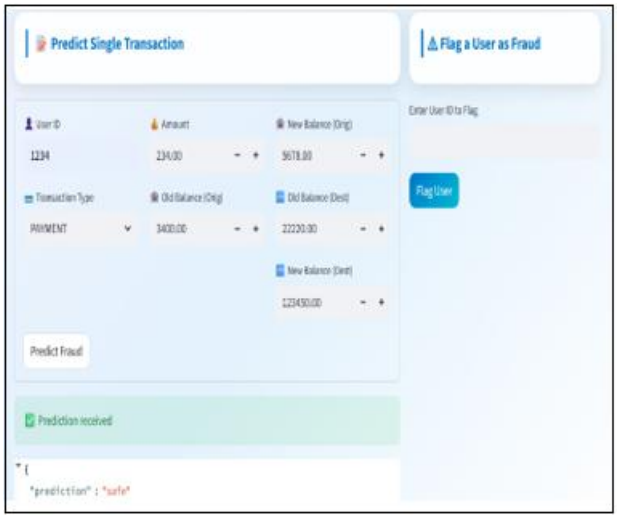

Figure 4: Prediction of Transaction

4. Conclusions

Key issues like data imbalance, changing fraud patterns, and the requirement for real-time analysis are all successfully addressed by the suggested AI-driven Financial Fraud Detection System. The system detects fraudulent transactions with high accuracy, scalability, and user engagement by integrating LightGBM into a Streamlit-based interface. While Explainable AI tools (SHAP and LIME) guarantee transparency and interpretability in model predictions, SMOTE improves the model's capacity to detect uncommon fraud cases.

The "Flagged as Fraud" feature, which automatically flags

suspicious transactions based on fraud probability scores, is a key feature of the system. Analysts can examine and take immediate action to stop losses thanks to the dashboard's safe storage and display of these flagged records. All things considered, this project offers a reliable, explainable, and real-time financial fraud detection solution, enabling organizations to lower risks, make wise decisions, and increase client confidence.

Acknowledgement
We would like to express our sincere gratitude to Prof. Swati Bagul. For their continuous support and guidance throughout the project. We also extend our thanks to JSPM'S Jayawantrao Sawant College of Engineering, Hadapsar, Pune for providing us with the necessary resources and infrastructure to carry out this research. Finally, we appreciate the valuable feedback from our project coordinator Prof. Swati Bagul, which helped improve the quality of this work.

References

[1] "Improved LightGBM for Extremely Imbalanced Data and Application to Credit Card Fraud Detection," IEEE Access, vol. 1, 2024, by X. Zhao, Y. Liu, and Q. Zhao. [Online].
[2] Y. Xie, G. Liu, C. Yan, C. Jiang, and M. Zhou, "Time-Aware Attention-Based Gated Network for Credit Card Fraud Detection by Extracting Transactional Behaviors," IEEE Transactions on Computational Social Systems, vol. 10, no. 3, pp. 1004–1016, 2022.
[3] "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 8, pp. 3784–3797, 2023; A. D. Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi
[4] B. Fetaji, M. Fetaji, A. Hasan, S. Rexhepi, and G. Armenski, "FRAUD-X: An Integrated AI, Blockchain, and Cybersecurity Framework with Early Warning Systems for Mitigating Online Financial Fraud — A Case Study from North Macedonia," IEEE Access, vol. 1, 2025.
[5] Usman, S. B., Abdullahi, Y. Liping, B. Alghofaily, A. S. Almasoud, and A. Rehman, "Financial Fraud Detection Using Value-at-Risk with Machine Learning in Skewed Data," IEEE Access, vol. 12, pp. 64285–64299, 2024.
[6] C. V. Sai, D. Das, N. Elmitwally, O. Elezaj, and M. Islam, "Explainable AI-Driven Financial Transaction Fraud Detection Using Machine Learning and Deep Neural Networks," SSRN, April 2024.
[7] K. Li, T. Yang, M. Zhou, J. Meng, S. Wang, Y. Wu, B. Tan, H. Song, L. Pan, F. Yu, Z. Sheng, and Y. Tong, "SEFraud: Graph-based Self-Explainable Fraud Detection via Interpretative Mask Learning," arXiv preprint, June 2024.