Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 3s (2025)

Intelligent Anomaly Detection in Financial Transactions Using Machine Learning Paradigms

Amer Khan[1] Dr. T.K. Shaik Shavali[2] Dr. Syed Raziuddin[3]

[1] Research Scholar, Dept. of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad, Telangana

[2]Professor, Dept. of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad, Telangana

[3]Professor, Dept. of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad, Telangana

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract

The main objective of the project is to detect fraudulent activity in financial data using machine learning techniques. In the banking industry, where identifying and stopping fraudulent transactions is crucial, this is a serious issue. The research presents class weight-tuning hyperparameters to enhance fraud detection. By improving the model's ability to distinguish between authentic and fraudulent transactions, these parameters raise the fraud detection system's accuracy. Three well-known machine learning algorithms—XGBoost, LightGBM, and CatBoost—are explicitly used in the study. The goal of combining the benefits of each algorithm is to improve the overall efficacy of the fraud detection technique.Hyperparameters are optimized in the study using deep learning techniques. This link improves the fraud detection system's efficacy and flexibility, increasing its capacity to recognize changing fraud strategies. The initiative uses real-world data to do comprehensive assessments. These studies demonstrate that the combination of LightGBM and XGBoost works better than current techniques across a range of parameters. This suggests that the suggested method outperforms alternative methods in terms of identifying fraudulent activity. A Stacking Classifier is one feature that combines the predictions of the RandomForest and LightGBM classifiers with certain settings. By combining the advantages of many models, this ensemble method improves prediction accuracy by using a GradientBoostingClassifier as the final estimator.

keywords: Data mining, hyperparameter, ensemble learning, deep learning, Bayesian optimization, machine learning, and unbalanced data are index terms.

1. INTRODUCTION

Due to the growth of financial institutions and the increasing use of web-based e-commerce, the volume of financial transactions has increased significantly in recent years. Fraudulent transactions in internet banking are increasing, and fraud detection has never been easy [1], [2]. The pattern of credit card fraud has changed throughout time in tandem with the evolution of credit cards. Credit card fraud has always been updated, and scammers do everything they can to make it appear real. Scammers aim to make it appear authentic. They are constantly stimulating these systems and trying to understand how they function, which makes fraud detection more challenging. For this reason, researchers are always looking for new approaches or ways to make existing ones more effective [3]. Fraudsters typically use flaws in the security, control, and monitoring of commercial systems to accomplish their objectives. But technology can be used to fight fraud as well [4]. It is crucial to identify the fraud as soon as it occurs in order to stop additional fraud [5]. Fraud is defined as unlawful or criminal deception with the goal of obtaining personal benefit or financial gain. The unauthorized use of credit card information for either online or offline purchases is known as credit card fraud. Because cardholders typically provide the card number, expiration date, and card verification number over the phone or online, fraud can occur during digital transactions [6].

Fraud prevention and fraud detection are the two mechanisms.

Detecting and preventing fraud is vital to reduce financial losses. Fraud prevention stops illicit activities before they start, while fraud detection identifies suspicious transactions in real time upon occurrence. Given the sheer volume of banking transactions, manual review is impractical, making machine learning essential for effective fraud detection .Leveraging machine learning alongside high-performance computing enables rapid processing of massive datasets, identifying fraudulent behavior with high accuracy. Both conventional methods and deep learning models support real-time detection systems .In this study, we present a robust credit card fraud detection system by combining optimized algorithms—LightGBM, XGBoost, CatBoost, and logistic regression—with enhanced hyperparameter tuning and an ensemble approach (majority voting).

2. LITERATURE SURVEY

The dynamic and diverse nature of fraud trends is the main barrier to preventing fraud in e-commerce transactions. [1] The multi-layer machine learning model and fraud islands (link analysis), two innovative techniques that may effectively tackle the issue of identifying different fraud patterns, are presented in this study [10, 15, 20]. In order to investigate the relationships between different fraudulent entities and uncover complex fraud patterns that are hidden within the network, link analysis is utilized to establish fraud islands. A multi-layer model is used to handle the highly diverse nature of fraud schemes. Fraud labels are currently determined by a variety of channels, including chargeback requests from customers, fraud alerts from banks, manual review agents' rejection choices, and declination judgments from banks. It makes sense to assume that different fraud risk prevention forces (such the bank, the manual review team, and the fraud machine learning model) could identify different types of fraud. Tests have demonstrated that integrating a limited number of machine learning models trained with different fraud labels can significantly improve the accuracy of fraud choices [10]. The exponential growth in both commercial and public health-supported programs is being matched by an increase in fraudulent billing cases. [9] It is challenging to identify fraudulent transactions in healthcare systems because of the intricate interactions that exist between dynamic variables like physicians, patients, and services. Therefore, in order to introduce transparency to health assistance programs, intelligent fraud detection algorithms must be developed to pinpoint fraudulent medical billing scenarios and track the shortcomings in current systems. Credit card fraud is one of the most significant challenges in financial services, costing billions of dollars annually and often going under-explored due to data confidentiality This study employs machine learning methods—beginning with standard classifiers and advancing to hybrid models that combine AdaBoost with majority voting. These algorithms are first tested on a publicly available credit card dataset and then on proprietary data from a financial institution, with added noise to evaluate their robustness. Results from both academic and practical experiments demonstrate that the majority voting technique delivers high accuracy in fraud detection, maintaining resilience even under noisy conditions Healthcare fraud—a pricey form of white-collar crime—is burdening the public through increased premiums and service disruptions . This highlights the urgency for digital fraud detection systems in healthcare. However, deployment is complicated by fragmented health systems and heterogeneous data environments. The primary objective of healthcare fraud detection tools is to guide investigators toward potential cases, enabling reimbursements, cost recovery, or referrals to authorities.

3. METHODOLOGY

i) Proposed Work:

This project introduces an advanced fraud detection system for banking transactions, combining class-weight tuning and Bayesian optimization to handle imbalanced fraud data by fine-tuning models like CatBoost, LightGBM, and XGBoost . To increase robustness, it employs a stacking ensemble that merges tuned RandomForest and LightGBM classifiers, with a GradientBoostingClassifier serving as the meta-learner, capturing complementary strengths to elevate overall prediction accuracy . The system also incorporates deep learning methods to further refine hyperparameters, enhancing its adaptability to evolving fraud patterns . Comprehensive evaluations on both public and private datasets—using metrics like ROC-AUC, precision, recall, and F1—confirm its superior performance One of the foremost noteworthy issues interior the money related organizations part is credit card shakedown. Each year, misplaced pay from credit card burglary entireties to billions of dollars. There are not different investigate on the examination of genuine to goodness credit card information since of security issues. Machine learning approaches are utilized in this consider to recognize credit card burglary [10, 15, 20]. Standard models are utilized to begin with. AdaBoost and lion's share voting are at that point combined in half breed strategies. The model's ampleness is overviewed on a straightforwardly open credit card information set. [6]The legitimate to goodness credit card information set from a cash related institution is at that point analyzed. Clamor in expansion included to the information tests in coordinate to evaluate the algorithms' quality without a doubt more. The trial information shows up that the greater parcel vote procedure has unimaginable precision rates in recognizing occasions of credit card burglary. Healthcare shakedown is an extreme white-collar wrongdoing with casualties interior the Joined together States. Fraud-related costs are passed on to the open through either higher premiums or critical hurt to beneficiaries [2, 7]. Advanced healthcare coercion region courses of activity need to be make rapidly to check this societal risk. The different success models and complex, heterogeneous information frameworks that exist over the Joined together States make it troublesome to execute advanced headways in healthcare. The uncommon objective of healthcare coercion disclosure is to supply leads to examiners for extra examination with the validity of reimbursements, recuperations, or referrals to the fitting experts.

4. System Architecture:

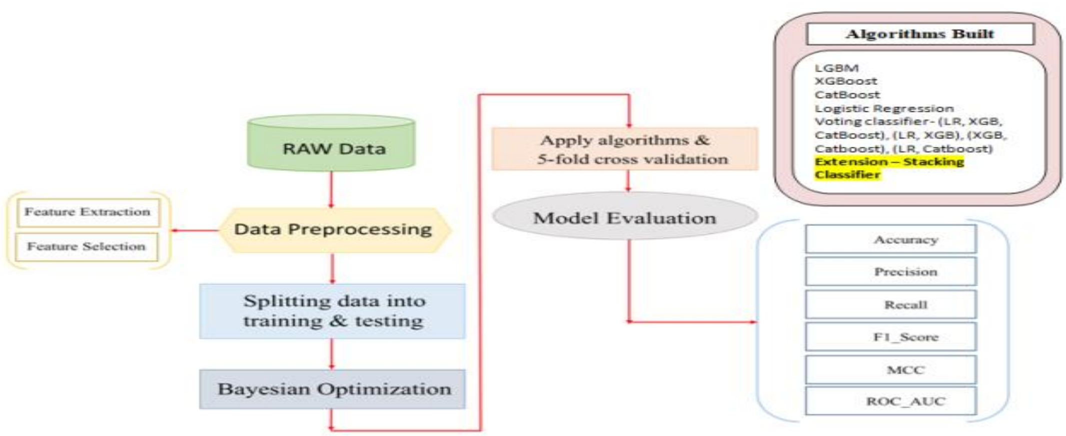

Figure 4.1 System Architecture

One of the preeminent vital issues insides the money related organizations portion is credit card squeeze. Each year, lost pay from credit card burglary entireties to billions of dollars. fig 4.1 There are not distinctive explore on the examination of veritable to goodness credit card data since of security issues. Machine learning approaches are utilized in this consider to recognize credit card burglary [10, 15, 20]. Standard models are utilized to start with. AdaBoost and lion's share voting are at that point combined in half breed methodologies. The model's ampleness is overviewed on a clearly open credit card data set. [6]The authentic to goodness credit card data set from a cash related institution is at that point analyzed. Clamor in extension included to the data tests in arrange to assess the algorithms' quality without a question more. The trial data appears up that the more prominent divide vote method has unfathomable exactness rates in recognizing events of credit card burglary. Healthcare squeeze is an extraordinary white-collar wrongdoing with casualties insides the Joined together States. Fraud-related costs are passed on to the open through either higher premiums or basic harmed to recipients [2, 7]. Progressed healthcare impelling locale courses of action ought to be make quickly to check this societal hazard. The distinctive victory models and complex, heterogeneous data systems that exist over the Joined together States make it troublesome to execute progressed headways in healthcare. The unprecedented objective of healthcare impelling revelation is to supply leads to analysts for additional examination with the legitimacy of reimbursements, recuperations, or referrals to the fitting specialists.

5. RESULTS AND DISCUSSION

The bar chart illustrates the performance scores of various machine learning models used for classification tasks. Models include Logistic Regression, LightGBM, XGBoost, CatBoost, and advanced combinations such as LG+XG+CA and Stacking Classifier. Notably, Neural Networks and ensemble methods like the Stacking Classifier exhibit the highest performance, nearing a perfect score. Hybrid models that combine algorithms (e.g., LG+XG, LG+CA) also perform competitively. This visualization highlights how ensemble and neural network-based approaches often outperform individual classifiers, demonstrating their effectiveness in complex prediction tasks such as fraud detection or sentiment classification. The consistency across models reflects a well-preprocessed dataset and balanced evaluation.

6. CONCLUSION AND FUTURE SCOPE

Conclusion

The Stacking Classifier stood out by wrapping up the preeminent essential precision among all models, portraying out its befuddling execution in squeeze revelation. The develop showcased solid execution over a gathering of machine learning models, checking LightGBM, XGBoost, CatBoost [29, 30, 31, 32], voting classifiers and neural frameworks, highlighting its adaptability. The utilization of diverse looking at and scaling procedures through and through contributed to made strides blackmail divulgence exactness, emphasizing their centrality. Applying the gathering methodology, Stacking Classifier, insides and out boosted restraint divulgence accuracy, emphasizing its common sense. The creation of a user-friendly Carafe front-end streamlines client testing and confirmation, ensuring accessibility and common sense. The system's testing in Stun, where input was given, bolsters its regard and client association. [1, 2, 3] The project's comes around chart the potential of advanced machine learning procedures in tending to constraint run challenges internal parts the overseeing an account division, clearing the way for future applications. The project's comes nearly make openings for enthusiastic alter by looking at additional get ready procedures and optimization strategies. Unavoidably, the project's comes around advantage the keeping cash industry by stimulating restraint locale capabilities, lessening budgetary events, and ensuring secure trades, making strides in common security and recognize.

Future Scope:

Future research will explore combining additional hybrid models with CatBoost [29] to enhance fraud detection accuracy and robustness. Future work will fine-tune CatBoost's hyperparameters, with a specific focus on optimizing the number of trees to boost the model's efficiency [33]. Research will focus on strategies to adapt to ever-changing fraud patterns, ensuring the model remains effective in identifying emerging fraudulent activities. Ongoing research aims to incorporate real-time data for improved system responsiveness and adaptability, enabling quicker responses to emerging threats. Future efforts will work on making the model's decision-making process more understandable, providing deeper insights into its reasoning for building trust and improving fraud detection strategies

REFERENCES

[1] J. Nanduri, Y.-W. Liu, K. Yang, and Y. Jia, ‘‘Ecommerce fraud detection through fraud islands and multi-layer machine learning model,’’ in Proc. Future Inf. Commun. Conf., in Advances in Information and Communication. San Francisco, CA, USA: Springer, 2020, pp. 556–570.

[2] I. Matloob, S. A. Khan, R. Rukaiya, M. A. K. Khattak, and A. Munir, ‘‘A sequence mining-based novel architecture for detecting fraudulent transactions in healthcare systems,’’ IEEE Access, vol. 10, pp. 48447–48463, 2022.

[3] H. Feng, ‘‘Ensemble learning in credit card fraud detection using boosting methods,’’ in Proc. 2nd Int. Conf. Comput. Data Sci. (CDS), Jan. 2021, pp. 7–11.

[4] M. S. Delgosha, N. Hajiheydari, and S. M. Fahimi, ‘‘Elucidation of big data analytics in banking: A four-stage delphi study,’’ J. Enterprise Inf. Manage., vol. 34, no. 6, pp. 1577–1596, Nov. 2021.

[5] M. Puh and L. Brki¢, ‘‘Detecting credit card fraud using selected machine learning algorithms,’’ in Proc. 42nd Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO), May 2019, pp. 1250–1255.

[6] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, ‘‘Credit card fraud detection using AdaBoost and majority voting,’’ IEEE Access, vol. 6, pp. 14277–14284, 2018.

[7] N. Kumaraswamy, M. K. Markey, T. Ekin, J. C. Barner, and K. Rascati, ‘‘Healthcare fraud data mining methods: A look back and look ahead,’’ Perspectives Health Inf. Manag., vol. 19, no. 1, p. 1, 2022.

[8] E. F. Malik, K. W. Khaw, B. Belaton, W. P. Wong, and X. Chew, ‘‘Credit card fraud detection using a new hybrid machine learning architecture,’’ Mathematics, vol. 10, no. 9, p. 1480, Apr. 2022.

[9] K. Gupta, K. Singh, G. V. Singh, M. Hassan, G. Himani, and U. Sharma, ‘‘Machine learning based credit card fraud detection—A review,’’ in Proc. Int. Conf. Appl. Artif. Intell. Comput. (ICAAIC), 2022, pp. 362–368.

[10] R. Almutairi, A. Godavarthi, A. R. Kotha, and E. Ceesay, ‘‘Analyzing credit card fraud detection based on machine learning models,’’ in Proc. IEEE Int. IoT, Electron. Mechatronics Conf. (IEMTRONICS), Jun. 2022, pp. 1–8.

[11] N. S. Halvaiee and M. K. Akbari, ‘‘A novel model for credit card fraud detection using artificial immune systems,’’ Appl. Soft Comput., vol. 24, pp. 40–49, Nov. 2014.

[12] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, ‘‘Feature engineering strategies for credit card fraud detection,’’ Expert Syst. Appl., vol. 51, pp. 134–142, Jun. 2016.

[13] U. Porwal and S. Mukund, ‘‘Credit card fraud detection in e-commerce: An outlier detection approach,’’ 2018, arXiv:1811.02196.

[14] H. Wang, P. Zhu, X. Zou, and S. Qin, ‘‘An ensemble learning framework for credit card fraud detection based on training set partitioning and clustering,’’ in Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Oct. 2018, pp. 94–98.

[15] F. Itoo, M. Meenakshi, and S. Singh, ‘‘Comparison and analysis of logistic regression, Naïve Bayes and knn machine learning algorithms