E-ISSN No : 2581-6195 | Volume : 9 | Special Issue | May 2026

FRAUD DETECTION IN BANKING DATA BY MACHINE LEARNING TECHNIQUES

Archana[1], Dr. Savita Soma[2]

Access Online

Article DOI:
10.5281/zenodo.21408930

[1] PG Student, E&CE (DCN) Dept. Guru Nanak Dev Engineering College, Bidar
[2] Assoc. Professor, E&CE Dept., Guru Nanak Dev Engineering College, Bidar (Karnataka)

How to Cite:
Archana & Dr. Savita Soma (2026), Fraud Detection in Banking Data by Machine Learning Techniques, International Educational Journal of Science & Engineering (IEJSE),Vol: 9, Special Issue, May 2026 206-210

ABSTRACT

Credit cards emerged as a prevalent payment method due To advance technological advances and expand e-commerce services. This has led to an increase in financial transactions. Furthermore, a significant increase in fraud has resulted in increased financial transaction fees. As a result, fraud research has proven to be a compelling topic. This study examines the use of class weight hyperparameters to adapt the weights of real and fraudulent transactions. Because real challenges such as unbalanced data are considered, Bayesian optimization is used to identify the optimal hyperparameter. We recommend Weight Voices as an initial measure of unbalanced data. DL is used to optimize the proposed method of hyperparameters, especially the weight division, to further improve performance. Empirical data are used to evaluate proposed strategies in the experimental environment. To handle imbalanced data records more effectively, "we use recall metrics in addition to traditional ROC-AUC". Evaluate CatBoost, LightGBM, XGBoost, and logistic regression separately using five cross-validation techniques. The performance of the combined algorithm is evaluated using the majority voice ensemble learning method. The results reveal that the recommended strategies work better than the most advanced ones and make a big difference in performance.

KEYWORDS: Fraud detection, Banking system, "ML and DL Models like LightGBM, XG Boost, Cat Boost, Neural Network and Hybrid model like LG + XG+ CAT, LG + XG, LG

INTRODUCTION

The The number of financial transactions has gone up a lot in the last several years since extra banks and credit unions have opened up and online shopping has become more popular. Fraudulent transactions in internet banking are more prevalent, although their detection has consistently posed challenges. The dynamics of credit card fraud have consistently evolved in tandem with the advancement of credit cards. An effective fraud detection system must identify a greater number of fraudulent cases while maintaining high accuracy, ensuring that all outcomes are correct. This will enhance client faith in the bank, and the institution will mitigate financial losses due to false positives. This article addresses the issue of fraudulent transactions in online banking. numerous challenges exist in detecting fraud, highlighting the necessity for effective methodologies. "The study advocates for the utilization of ML techniques such as CatBoost, LightGBM, and XGBoost, alongside DL and hyperparameter optimization", to enhance fraud detection efficacy. We employ Bayesian optimization to detect fraud and recommend the initial usage of the weight-tuning hyperparameter to cope with the issue of imbalanced data. We recommend employing CatBoost and unbalanced datasets. Furthermore, we compare overall performance the usage of F1 score and Matthew's correlation coefficient (MCC)" metrics. The consequences reveal that the proposed procedures . outperform the mounted methods. We make use of publicly reachable datasets, and extra students have to put up public get entry to supply codes.

Aim: The aim is to develop a machine learning-based model that accurately detects fraudulent banking transactions by analyzing historical transaction data, thereby improving the security and reliability of financial systems while minimizing financial losses and false alarms. The fraudulent detection in online banking is done by using multiple kinds of ML and DL models, such as LightGBM, XG boost, Cat boost, and Neural network. Hybrid models like LG + XG+ CAT, LG + XG, LG + CAT, XG + CAT.

Objective: With the rise of digital banking and online financial transactions, fraud in banking systems has become increasingly sophisticated and frequent. Detecting fraudulent activities such as unauthorized transactions, identity theft and account takeovers in real time poses significant challenges due to the high volume of data, class imbalance (very few fraud cases compared to legitimate ones), and the constantly evolving nature of fraudulent behavior. The objective of this project is to develop a machine learning-based system that can accurately detect fraudulent transactions in banking data. The system must be capable of learning patterns from historical transaction data and generalizing to identify potential fraud in new unseen data. It should XGBoost in conjunction with LightGBM to enhance performance. The XGBoost technique is employed due to its rapid training talents on large datasets and its incorporation of regularization, which mitigates overfitting via assessing model complexity. The construction of the tree and hyperparameters is expeditious. CatBoost is an additional algorithm that we utilize. As it is superfluous to modify the hyperparameters for overfitting management, it also yields effective results. without altering the hyperparameters in relation to other algorithmic learning methodologies. "Integrate CatBoost, XGBoost, and LightGBM methodologies and assess the effect of these combined strategies" on fraud detection performance using real, unbalanced data. Furthermore, we recommend employing DL for the modification of hyperparameters, among other adjustments. To evaluate the efficacy of the proposed methods, we offer complete critiques making use of empirical data. along with the frequently applied ROC- AUC, we additionally include remember precision to extra correctly deal with address issues such as data imbalance, feature selection, and real-time detection requirements while minimizing false positives and false negatives.

MATERIALS AND METHODS

The Materials and Methods section should be brief and concise with sufficient technical information. Only new methods should be described in detail. Cite previously published procedures in References.

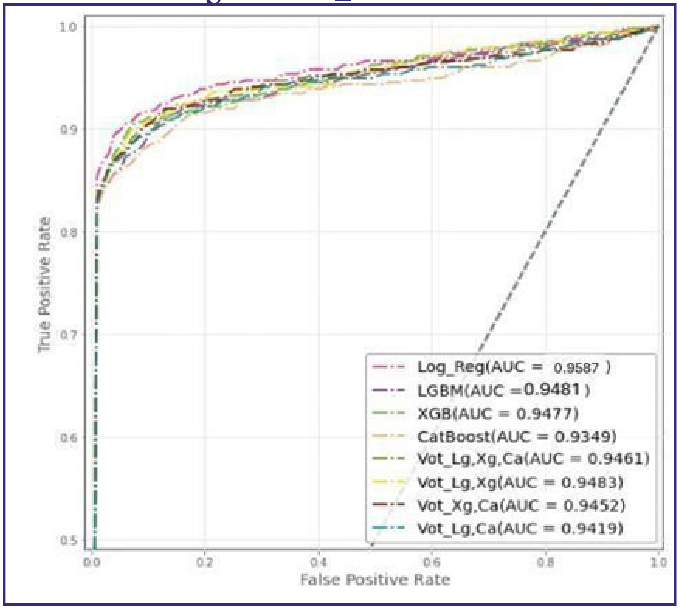
RESULTS

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. We examine the algorithms in triple and double precision. The comparison results are presented in table 1.

Model	Accuracy	AUC	Recall	Precision	F1-score	MCC
Log_Reg	0.97474	0.9587	0.8733	0.7361	0.1145	0.2246
LGBM	0.99914	0.9481	0.7992	0.7539	0.7696	0.7724
XGB	0.99922	0.9519	0.7951	0.7363	0.7832	0.7867
CatBoost	0.99878	0.9393	0.8098	0.7673	0.7069	0.7159
Vot_ Lg,Xg, Ca	0.99926	0.9502	0.8035	0.7228	0.7829	0.7844
Vot_Lg, Xg	0.99928	0.9525	0.8013	0.7326	0.7903	0.7926
Vot_g, Ca	0.99924	0.9493	0.8099	0.7613	0.7824	0.7856
Vot_Lg, Ca	0.99914	0.9456	0.8076	0.7262	0.7583	0.7623

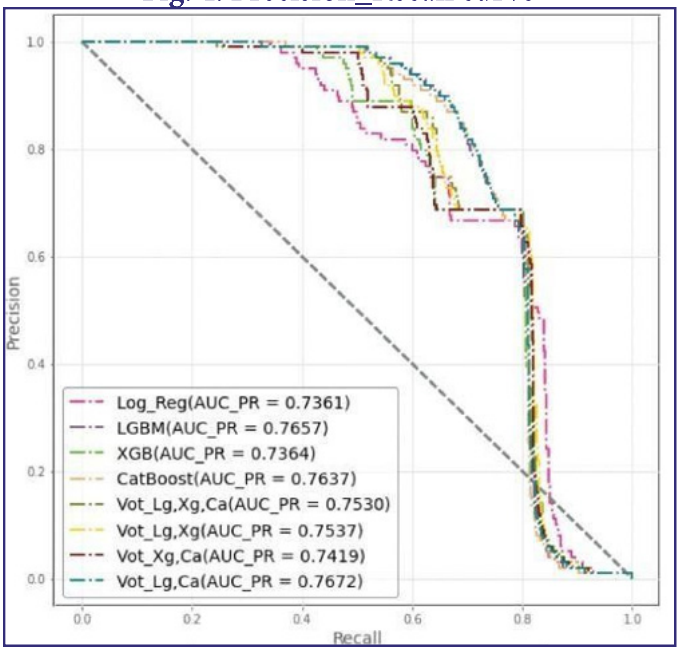
Most studies in the literature rely on AUC diagrams to evaluate performance. However, as can be seen from the ROC-AUC curve in Fig. 3, the value of AUC in severely unbalanced data is not a good evaluation metric. It is influenced by the real positives and considers the negatives irrelevant. According to the ROC-AUC Fig. 3, the logistic regression algorithm 0.9587 has the highest number of fraud detection. but it has the lowest value in other criteria.

Fig. 3: ROC_AUC curve



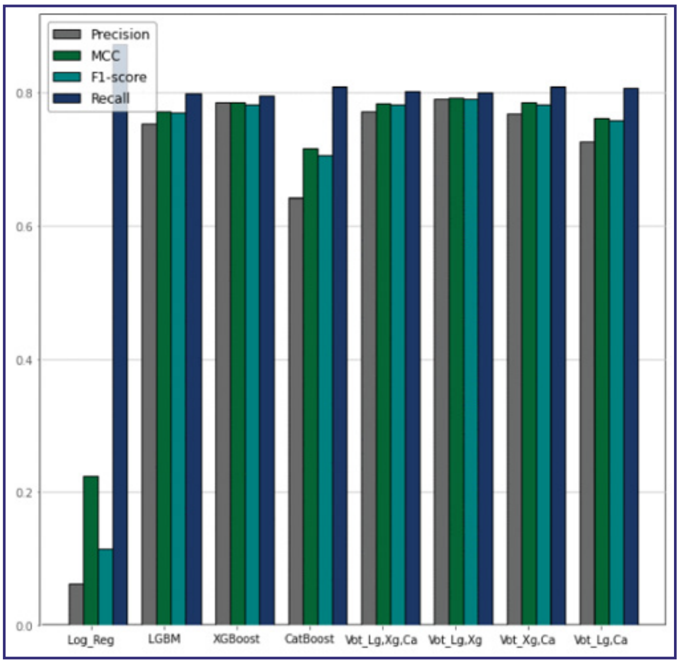
The precision-recall curve is illustrated in Fig.4 and shows the system performance in a more precise manner compared with the ROC-AUC curve. The highest value belongs to the combination of the CatBoost and LightGBM algorithms with a value of 0.7637, and the lowest value belongs to logistic regression and is 0.7361.

Fig. 4: Precision_Recall curve



Comparing the precision, recall, and F1-score as well as the MCC, the algorithms used are shown in Fig.5. The best performance is related to the combination of lightGBM and XGBoost algorithms, which have an MCC value of 0.79 and F1-score of 0.79. In individual algorithms, XGBoost has the highest values.

Fig. 5: Performance comparing algorithms with different evaluation criteria



DISCUSSION

The precision-recall curve is illustrated in Fig.4 and shows the system performance in a more precise manner compared with the ROC-AUC curve. The highest value belongs to the combination of the CatBoost and LightGBM algorithms with a value of 0.7637, and the lowest value belongs to logistic regression and is

CONCLUSIONS

The conclusion of the paper is that the use of Cat-Boost, LightGBM, XG-BOOST and DL of fine dollars of hyperparameters to significantly improve the performance of fraud detection in real unmistakable data records. This proposed method can find false transactions in banking data and reduce the high cost of banking transactions related to fraud. The results show that the recommended strategies work better than the others and change the big difference in performance. This research recommends that alternative hybrid models can be used for future work, and more hyperparameters can be changed in the Cat Boost field, especially the number of trees that can improve the performance of this proposed model. The fact that MCC can provide reliable results for heterogeneous data has shown that it is stronger than other evaluation criteria. We got 0.79 and 0.81 for the DL method by using both the LightGBM and XG boost methods in this research. Using hyperparameters to fix data imbalance is preferable than using sample methods since it Speeds up the system of evaluating algorithms and uses less memory. We endorse adopting greater hybrid fashions and running especially withinside the place of Cat raises through adjusting greater hyper parameters, especially the quantity of bushes hyper parameter, for destiny research and development. Also, due to the hardware regulations on this investigation, the usage of higher and more potent hardware can also additionally cause advanced consequences that may be as compared to the consequences of this study..

ACKNOWLEDGMENTS

The authors thanks for their help in reviewing this manuscript.

TABLES AND FIGURES

All the tables and figures should be in the text at suitable place. Tables must be created using Word table format.

Tables must have titles and sufficient empirical detail in a legend immediately following the title to be understandable without reference to the text. Each column in a table must have a heading, and abbreviations, when necessary, should be defined in the legend. Please number the tables.

Figures should have titles and explanatory legends containing sufficient detail to make the figure easily understood.

Appropriately sized numbers, letters, and symbols should be used. The abscissa and ordinate should be clearly labeled with appropriately sized type.

Images or Diagrams should be very clear and high resolution. Blur or tiny images may not accept.

REFERENCES AND FOOTNOTES

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. Brkić, "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/S CI), 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 for credit card fraud detection," Int. J. Inf. Technol., vol. 13, no. 4, pp. 1503–1511, 2021.
16. T. A. Olowookere and O. S. Adewale, "A framework for detecting credit card fraud with cost- sensitive meta-learning ensemble approach," Sci. Afr., vol. 8, Jul. 2020, Art. no. e00464.

17. A. A. Taha and S. J. Malebary, "An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine," IEEE Access, vol. 8, pp. 25579–25587, 2020.
18. X. Kewei, B. Peng, Y. Jiang, and T. Lu, "A hybrid deep learning model for online fraud detection," in Proc. IEEE Int. Conf. Consum. Electron. Comput. Eng. (ICCECE), Jan. 2021, pp. 431–434.
19. T. Vairam, S. Sarathambekai, S. Bhavadharani, A. K. Dharshini, N. N. Sri, and T. Sen, "Evaluation of Naïve Bayes and voting classifier algorithm for credit card fraud detection," in Proc. 8th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), Mar. 2022, pp. 602–608.
20. P. Verma and P. Tyagi, "Analysis of supervised machine learning algorithms in the context of fraud detection," ECS Trans., vol. 107, no. 1, p. 7189, 2022.
21. Rahul Roy Devarakonda, "Machine Learning Approach for Fraud Detection in a Financial Services Application", E-ISSN: 2229-7677. 2023
22. Sudheer Chennuri, Advancing Fraud Detection in Banking: Integration of Data Pipelines, Machine Learning, and Cloud Computing", E-ISSN: 2582- 2160, 2024.
23. Yisong Chen, Chuqing Zhao, "Year over year developments in financial fraud detection by machine learning: A systematic literature survey", 2502.00201v1, 2025.