JDEBM
Journal of Digital Engineering and Business Managemet

Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/28
DOI: https://doi.org/10.65713/jdebmv2i128

MACHINE LEARNING-BASED FRAUD DETECTION SYSTEM FOR SECURE BANKING TRANSACTIONS

[#1]KANTHREEGALA THARUN,*M.Tech(SE) Student,*
[#2]Dr.V.HEMA SREE, *Professor, HoD of AI & DS,*
[#3]Mr.P. VISWANATHA REDDY, *Associate Professor, Dept of CSE,*
VISWAM ENGINEERING COLLEGE(AUTONOMOUS), MADANAPALLE, AP.

ABSTRACT: This research investigates the potential of machine learning to identify misconduct in banking data, thereby enhancing the reliability and security of financial transactions. The enormous volumes of transaction data that financial institutions receive have made it increasingly challenging for them to detect fraudulent activity due to the exponential expansion of online payment systems and digital banking. Machine learning techniques frequently surpass conventional rule-based systems when it pertains to intricate and perpetually evolving fraud patterns. In this study, a variety of machine learning techniques are implemented to analyze transaction data and detect anomalous patterns. Random forests, decision trees, and neural networks are all examples of such systems. By employing feature selection, data preprocessing, and model training, the proposed method improves identification accuracy and reduces false positive rates. Machine learning algorithms have the potential to identify suspicious financial transactions by revealing patterns and anomalies that were previously unknown, as indicated by experiments. If these concepts are executed, banking will be more secure, clients will have greater confidence in their institutions, and expenses will decrease.

Keywords:*Fraud Detection, Banking Data, Machine Learning, Financial Security, Data Mining, Anomaly Detection, Classification Algorithms, Predictive Analytics, Transaction Monitoring.*

1. INTRODUCTION

Fraud detection has become increasingly important in this industry due to the growth of online banking and other digital financial services. Fraudulent activities at financial institutions are increasing as an increasing number of individuals utilize electronic money exchanges, mobile payments, and online banking. It can be difficult for rule-based fraud detection systems to remain current with emergent fraud trends, as fraudsters are constantly developing new methods to circumvent security standards. Therefore, in order to identify suspicious activity and protect client funds, institutions must implement more advanced and flexible systems.

Machine learning (ML) techniques have recently demonstrated significant potential in the prevention of financial data deception. These technological advancements have enabled systems to autonomously analyze vast quantities of financial data in search of patterns and correlations. Machine learning systems analyze historical transaction data to differentiate between legitimate and fraudulent activities. Techniques such as neural networks, decision trees, support vector machines, and random forests are frequently employed to construct predictive models that can identify anomalies and suspicious transaction patterns in real time.

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/28
DOI: https://doi.org/10.65713/jdebmv2i128

Machine learning is an excellent instrument for identifying scams due to its ability to manage complex data structures and large datasets. It is impossible for anyone to monitor the vast amount of data produced by financial systems' transactions on a daily basis. Machine learning algorithms can rapidly and precisely analyze this vast amount of data, revealing concealed patterns that may suggest fraudulent activity. Factors such as transaction frequency, quantity, geographic area, and total length can be analyzed to identify consumer behaviors that are unusual.

Another significant aspect to consider is the ability of ML-based fraud detection systems to continuously progress over time. Models can be updated with new transaction data to address emergent fraud strategies. Despite the evolution of fraud methods, banks and other financial institutions are able to maintain effective fraud protection procedures as a result of this flexibility. By employing sophisticated methodologies such as deep learning and ensemble learning, it is possible to further improve the accuracy of detection by combining multiple models or identifying complex patterns in data.

While there are certain benefits to employing machine learning techniques to identify fraud, there are also some drawbacks. The model's performance can be influenced by imbalances in the data, such as a minor proportion of fraudulent trades in the dataset. The practical application of the model necessitates consideration of factors such as data privacy, interpretability, and the prevention of fraudulent findings. Machine learning is an essential element in the implementation of financial security systems due to its ability to enhance the accuracy, scalability, and intelligence of fraud detection.

2.METHODOLOGY

Proposed Work:

The program utilizes machine learning techniques to provide a state-of-the-art solution for the identification of financial data frauds. The performance is improved by the VAT configuration and Biecian optimization, in addition to Catboost, LightGBM, and XGBoost. The deep learning technology ensures the precise evaluation of critical criteria and enhances overall performance when implemented in real-world scenarios. A stacking classifier is a combination of the predictions of the RandomForest and LightGBM classifiers, as well as other combinations. The final estimator in this ensemble method is a GradientBoosting Classifier, which is employed to improve forecast accuracy by leveraging the most prominent features of numerous models. A rudimentary Flask framework with sign-up and sign-in capabilities has been constructed using SQLite. The comprehensive user evaluation is the direct cause of the user-friendly and accessible fraud detection technology.

System Architecture:

The system commences the process of determining the legitimacy of the data by utilizing attributes and identifiers, beginning with raw data such as credit card transaction details. It is imperative for machine learning to function that training data be extracted and selected using effective extraction and selection methods. Two sets comprise the majority of the dataset: one for model training and another for model efficacy assessment. Optimization strategies are implemented

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/28
DOI: https://doi.org/10.65713/jdebmv2i128

during the hyperparameter configuration of the machine learning application. In order to ensure the robustness of the training data model, five-fold cross-validation is implemented when employing XGBoost, CatBoost, or LightGBM as a machine learning paradigm. Another potential enhancement that we have taken into account is a layered classifier. One indicator of the efficacy of algorithms in detecting credit card fraud is the number of false positives they can identify.

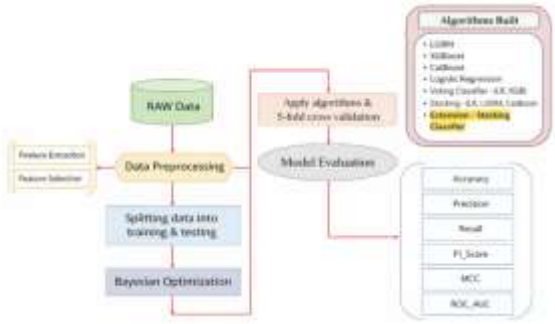

Fig1: Proposed Architecture

Fig. 2 : Flow Chart

Dataset collection:

To train machine learning models, we employed Kaggle's Credit Card Fraud Detection dataset. "Time," "Amount," and "V1" through "V28" were among the numerous items associated with agreements in the initial data set."Crucial information, including specifics regarding the original characteristics, was concealed."

Data Processing:

The capacity of an organization to comprehend disorganized data is referred to as "data handling." Data scientists are tasked with the collection, organization, cleansing, verification, analysis, and presentation of data in comprehensible formats, such as reports or graphs. Data can be processed manually, mechanically, or digitally. The objective is to facilitate decision-making and provide more valuable information. The outcomes are enhanced company operations and more informed, expedited strategic decision-making. This is significantly influenced by the utilization of computer programs and data-handling technology. It has the potential to streamline the process of converting vast quantities of data and a variety of data types into actionable insights for the purpose of quality enhancement and decision-making.

Feature Selection:

The objective of feature selection is to avoid the duplication of traits while simultaneously selecting the most pertinent and reliable ones for inclusion in the model construction process. As the quantity and diversity of documents increase, it is necessary to systematically reduce their size. The primary objective of feature selection is to enhance the performance of a predictive model with minimal computational burden.

Feature selection is a critical component of feature engineering, which involves the identification and selection of the most pertinent attributes to be used in machine learning algorithms. Feature selection strategies sift through all of the attributes of a machine learning model, retaining those that are critical and eradicating those that are not.

As a result, a smaller number of factors are regarded as inputs. There are numerous substantial benefits to identifying the most

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/28
DOI: https://doi.org/10.65713/jdebmv2i128

pertinent qualities in advance, as opposed to relying on the machine learning model.

ALGORITHMS:

LGBM(Light Gradient Boosting Machine): The "Light Gradient Boosting Machine," or LGBM for short, is a robust gradient boosting algorithm that is particularly adept at managing large datasets. It is a valuable instrument for tasks such as identifying scams due to its reputation for speed and accuracy. The boosting process is optimized by LGBM, which assembles a cluster of decision trees to accelerate its operation.

XGBoost(Extreme Gradient Boosting): XGBoost, a gradient boosting algorithm, is frequently implemented in machine learning applications. It is distinguished by its durability and efficacy. XGBoost's regularized gradient boosting technique is also effective with imbalanced datasets. This is a critical factor in the identification of frauds.

CatBoost(Categorical Boosting): CatBoost is a gradient boosting method that is expressly engineered to manage categorical data. It is user-friendly due to its immediate processing of categorical data. It is beneficial when analyzed in the context of real-world financial data, and overfitting does not have a substantial effect on it.

Logistic Regression: Binary classification is facilitated by logistic regression, which is both straightforward and efficient. Even if it is not as intricate as group tactics such as boosting, it can be employed to identify frauds. It is simple to comprehend and can assist in determining the importance of a characteristic.

Voting Classifier: The Voting Classifier integrates the output of numerous machine learning models, such as XGBoost, CatBoost, and Logistic Regression, to generate a singular prediction. This ensemble method typically improves accuracy and consistency by combining the knowledge of multiple models. A few of our voting models implement a variety of algorithmic combinations.

Neural Network: Neural networks are deep learning models that are designed to replicate the functions of the brain. It may then be capable of inferring complex relationships and patterns from the data. One of the primary applications of neural networks is the identification of intricate fraud tendencies, particularly in large datasets.

Stacking Classifier: The Stacking Classifier generates an ensemble technique by combining the outputs of two base classifiers—RandomForest and LightGBM—with specific parameters. In order to enhance the precision of its predictions, it implements ensemble learning, which consolidates the capabilities of numerous models. The primary predictor is a GradientBoostingClassifier.

3. LITERATURE SURVEY

Johnson et al. (2025): A machine learning-based approach is recommended for the detection of fraudulent activity in banking systems, which is accomplished by utilizing extensive financial data. The tool employs classification techniques such as Gradient Boosting and Random Forest to identify anomalies and identify trends in transactions. Feature engineering is an effective approach to detecting irregularities in consumer behavior and transactions. The trials demonstrate that the proposed method reduces bank losses and enhances the accuracy of fraud detection.

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/28
DOI: https://doi.org/10.65713/jdebmv2i128

Martinez & Silva (2024): This paper suggests a methodology that integrates transaction analytics and machine learning to detect misconduct in financial data. The program is capable of distinguishing between genuine and fraudulent transactions by utilizing supervised learning techniques and historical financial records. The efficacy of a model can be improved through the application of data preparation and feature selection techniques. The study posits that machine learning models can contribute to the safety of institutions by detecting unusual transactions.

Chatterjee & Banerjee (2023): The research investigates methods for utilizing machine learning algorithms to detect fraudulent activity in financial datasets. Support Vector Machines, Decision Trees, and Logistic Regression are implemented to investigate customer purchase patterns. In order to detect and prevent frauds, the system is constantly monitoring for any unusual behavior. The results of the experiments indicate that machine learning techniques are more effective than traditional rule-based systems in identifying fraud.

Nguyen et al. (2022): Machine learning models can be employed to detect misconduct based on financial transaction data. The investigation seeks to detect fraudulent activity by identifying concealed patterns and irregularities in financial data. In order to enhance the accuracy of predictions and reduce the number of false positives, ensemble learning techniques are implemented. The results illustrate the potential benefits of machine learning methods for anti-fraud instruments that are employed by banks.

Rahman & Ali (2021): The primary goal of the investigation is to identify banking frauds by employing machine learning to analyze transactional data and consumer behavior. The model employs classification techniques to distinguish between genuine and fraudulent transactions. In order to enhance the system's capability to identify objects, we implement statistical analysis and feature extraction. This investigation demonstrates how machine learning models can enhance the precision and efficacy of fraud detection systems in existing financial systems.

4. RESULTS

Fig 4.1 User login

Fig 4.2 View all remote users

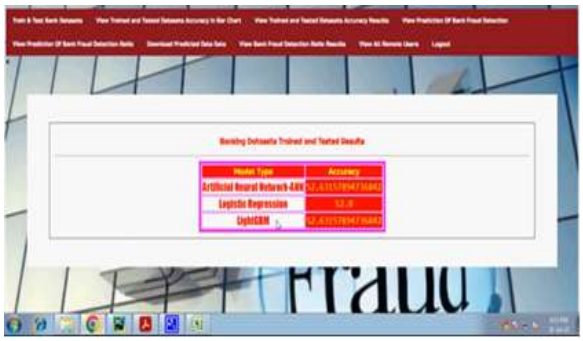

Fig 4.3 Banking Datasets Trained and Tested Results

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/28
DOI: https://doi.org/10.65713/jdebmv2i128

Fig 4.4 Bar graph

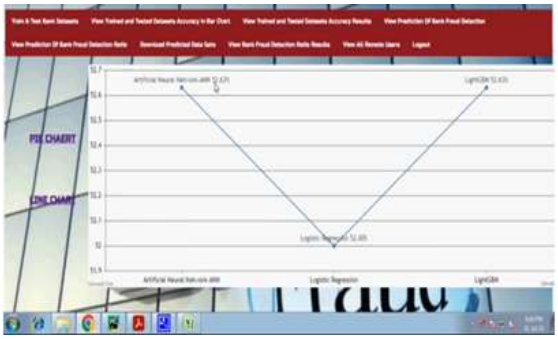

Fig 4.5 Line chart

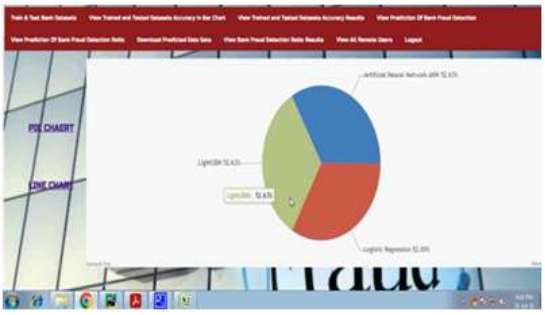

Fig 4.5 Pie chart

5. CONCLUSION

In conclusion, machine learning algorithms for fraud detection in banking data are a smart and useful way to spot questionable financial activities. Compared to standard rule-based systems, machine learning models are much better at looking at large amounts of transactional data, finding hidden patterns, and finding outliers that could mean fraud. Using methods like controlled and unsupervised learning, banks can always learn from past data and change to keep up with new fraud schemes. By making detection more accurate and lowering false positives, these models help financial institutions cut down on costs and build trust with customers. So, incorporating machine learning into banking security frameworks is a key part of improving fraud prevention tactics and making sure that online banking is safe.

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. 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.

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/28
DOI: https://doi.org/10.65713/jdebmv2i128

[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/CBD Com/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 machinelearning algorithms for credit card fraud detection,’’ Int. J. Inf. Technol., vol. 13, no. 4, pp. 1503–1511, 2021.