AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(1) (2026)
www.ajmimc.com

Financial Fraud Detection Using Value At Risk With Machine Learning In Skewed Data

[1]M. Madhavi Latha,[2]K.Varshitha,[3]P.Haripriya,[4]K.Shireesha,[5]CH.Sadhana,[6]R.Vennala

[1]Assistant Professor, Department of Computer Science & Cyber Security,

Princeton Institute of Engineering & Technology For Women

[2,3,4,5,6]B. Tech Students, Department of Computer Science & Cyber Security,

Princeton Institute of Engineering & Technology For Women

ABSTRACT

Financial fraud has become a major challenge for banking and financial institutions due to the rapid growth of digital transactions and online payment systems. Detecting fraudulent activities is particularly difficult when transaction datasets are highly imbalanced or skewed, where legitimate transactions significantly outnumber fraudulent ones. Traditional rule-based systems often fail to identify complex and evolving fraud patterns, making it necessary to adopt intelligent data-driven approaches. This study presents a machine learning-based financial fraud detection framework that incorporates the concept of Value at Risk (VaR) to improve the identification of suspicious transactions in skewed datasets. The proposed approach utilizes VaR as a financial risk indicator to estimate potential losses in transaction amounts and integrate it as an additional feature in the fraud detection process. By combining risk-based financial analysis with machine learning techniques, the system is able to capture abnormal transaction behaviors more effectively. The dataset undergoes preprocessing steps including data cleaning, feature selection, categorical encoding, and normalization to prepare the data for model training. Multiple machine learning algorithms such as Logistic Regression, Gradient Boosting, and XGBoost are employed to classify transactions as fraudulent or legitimate. Since fraud detection datasets are typically imbalanced, appropriate techniques such as class weighting and stratified data splitting are used to address skewed data distribution and improve model performance. The models are evaluated based on accuracy and predictive capability, and the best-performing model is selected for final fraud prediction. Experimental results demonstrate that incorporating Value at Risk as a feature enhances the detection capability of machine learning models in identifying risky transactions. The proposed framework provides an effective and scalable solution for financial fraud detection by integrating financial risk metrics with machine learning techniques. This approach can help financial institutions detect fraudulent transactions more accurately, reduce financial losses, and strengthen the security of digital financial systems.

Keywords: Financial Fraud Detection, Machine Learning, Value at Risk (VaR), Imbalanced Data, Skewed Datasets, Risk-Based Analysis, Predictive Analytics, Logistic Regression, Gradient Boosting, XGBoost, Data Preprocessing, Feature Engineering, Class Imbalance Handling, Anomaly Detection, Transaction Classification.

INTRODUCTION

The rapid expansion of digital banking, online payment platforms, and electronic financial services has significantly increased the volume of financial transactions conducted every day. While these technological advancements provide convenience and efficiency for customers and financial institutions, they also create new opportunities for fraudulent activities. Financial fraud, particularly transaction fraud, has become a major concern for banks, payment gateways, and e-commerce platforms, leading to substantial financial losses and security risks. Detecting fraudulent transactions accurately and at an early stage is therefore essential to ensure the safety and reliability of financial systems.

One of the major challenges in fraud detection is the skewed nature of transaction

datasets. In most financial datasets, legitimate transactions vastly outnumber fraudulent ones, resulting in a highly imbalanced or skewed data distribution. Traditional statistical or rule-based fraud detection methods often fail to perform effectively in such conditions because they may be biased toward the majority class. As a result, fraudulent transactions may remain undetected. To address this problem, machine learning techniques have been widely adopted, as they can analyze large datasets, identify hidden patterns, and detect anomalies that indicate fraudulent behavior.

Machine learning algorithms such as Logistic Regression, Gradient Boosting, Random Forest, and XGBoost have shown strong performance in financial fraud detection tasks. These models can learn from historical transaction data and classify new transactions as legitimate or fraudulent. However, improving the predictive performance of these models often requires the integration of meaningful financial risk indicators that capture unusual transaction behavior.

One such important financial risk measurement is Value at Risk (VaR). VaR is a widely used statistical technique in financial risk management that estimates the potential loss in the value of an asset or portfolio within a specified confidence level and time period. By incorporating VaR into the fraud detection process, it becomes possible to identify transactions that carry higher financial risk and may indicate suspicious activity. Using VaR as an additional feature can help machine learning models better distinguish between normal and abnormal transaction patterns, particularly in skewed datasets.

This study proposes a financial fraud detection framework that integrates Value at Risk with machine learning algorithms to handle skewed transaction data effectively. The system performs data preprocessing, feature engineering, and risk calculation to enhance the dataset before training multiple machine learning models. The trained models are then evaluated to determine the most effective approach for fraud classification. By combining financial risk analysis with machine learning techniques, the proposed framework aims to improve the accuracy and reliability of fraud detection systems.

Overall, this approach contributes to the development of intelligent fraud detection systems that can assist financial institutions in identifying suspicious transactions more efficiently, reducing financial losses, and strengthening the security of digital financial transactions.

I. LITERATURE SURVEY

1. Title: Credit Card Fraud Detection Using Machine Learning

Authors: Andrea Dal Pozzolo, Olivier Caelen, Reid A. Johnson, Gianluca Bontempi

Abstract:

This study presents a machine learning approach for detecting credit card fraud in highly imbalanced transaction datasets. The authors analyze the challenges associated with skewed data distributions where fraudulent transactions represent only a small portion of the dataset. Several machine learning techniques such as logistic regression, decision trees, and ensemble models are evaluated to improve fraud detection accuracy. The study emphasizes the importance of data preprocessing, feature engineering, and evaluation metrics suited for imbalanced datasets. Experimental results demonstrate that machine learning models can significantly improve the identification of fraudulent transactions compared to traditional rule-based systems.

2. Title: Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy

Authors: Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi, Gianluca Bontempi

Abstract:

This research focuses on developing realistic fraud detection models for credit card transactions. The authors address the issue of concept drift and class imbalance commonly found in fraud datasets. A novel learning strategy is proposed to handle highly skewed data distributions while maintaining high detection performance. The study highlights the need for adaptive learning models capable of detecting

evolving fraud patterns. Experimental analysis shows that the proposed method improves detection rates and reduces false alarms in real-world financial datasets.

3. Title: Financial Fraud Detection Using Machine Learning Techniques

Authors: J. West and M. Bhattacharya

Abstract:

This paper explores the application of machine learning algorithms in detecting financial fraud within banking systems. Various classification techniques including neural networks, decision trees, and support vector machines are analyzed for their effectiveness in fraud detection tasks. The study emphasizes the importance of feature selection and data preprocessing to enhance model performance. Results indicate that machine learning models can detect fraudulent transactions with higher accuracy compared to traditional detection methods.

4. Title: A Survey of Credit Card Fraud Detection Techniques

Authors: Clifton Phua, Vincent Lee, Kate Smith, Ross Gayler

Abstract:

This survey paper reviews various fraud detection techniques used in credit card transaction analysis. The authors categorize fraud detection methods into statistical approaches, machine learning techniques, and hybrid models. The study highlights key challenges such as data imbalance, evolving fraud strategies, and the need for real-time detection systems. The survey concludes that machine learning and hybrid approaches provide better performance in detecting fraudulent activities in large-scale financial datasets.

5. Title: Machine Learning for Financial Fraud Detection: A Review

Authors: S. Abdallah, M. Maarof, and A. Zainal

Abstract:

This paper presents a comprehensive review of machine learning methods applied to financial fraud detection. The study examines various algorithms such as logistic regression, decision trees, random forests, and gradient boosting methods. It also discusses the impact of imbalanced datasets on fraud detection performance and highlights the importance of feature engineering and risk analysis. The authors conclude that integrating financial risk indicators with machine learning models can significantly enhance fraud detection systems.

II. EXISTING SYSTEM

In traditional financial systems, fraud detection is mainly performed using rule-based monitoring and basic statistical analysis. These systems rely on predefined rules and thresholds to identify suspicious transactions, such as unusually high transaction amounts, multiple transactions within a short time period, or transactions from unfamiliar locations. Financial institutions typically maintain a set of fixed rules to flag potentially fraudulent activities and alert administrators for further investigation. Although such systems are simple to implement and easy to understand, they often fail to detect complex fraud patterns because fraudsters continuously change their techniques to bypass detection rules. Additionally, many existing systems do not effectively handle highly skewed datasets where legitimate transactions greatly outnumber fraudulent ones. As a result, traditional approaches struggle to identify rare fraudulent events and may generate inaccurate results or excessive false alerts.

III. PROPOSED SYSTEM

The proposed system introduces a machine learning-based financial fraud detection framework that integrates Value at Risk (VaR) to improve fraud identification in skewed transaction datasets. In this approach, financial transaction data is collected and preprocessed through steps such as data cleaning, removal of irrelevant attributes, categorical encoding, and feature scaling. A financial risk indicator, Value at Risk (VaR), is calculated from transaction amounts and incorporated as an additional feature to capture the potential financial risk associated with each transaction. The dataset is then divided into training and testing sets to ensure reliable model evaluation. Multiple machine learning algorithms such as Logistic Regression, Gradient Boosting, and XGBoost are trained

on the processed data to classify transactions as fraudulent or legitimate. Since fraud datasets are typically imbalanced, techniques such as class weighting and stratified splitting are applied to handle skewed data distribution. The models are evaluated based on performance metrics, and the best-performing model is selected for final fraud prediction. By combining financial risk analysis with machine learning techniques, the proposed system provides a more intelligent and adaptive approach for detecting fraudulent transactions and enhancing financial security.

IV. SYSTEM ARCHITECTURE

The system architecture illustrates the overall workflow of the financial fraud detection system that integrates Value at Risk (VaR) with machine learning techniques. Initially, raw transaction data is collected from financial records and passed to the preprocessing module, where data cleaning, categorical encoding, and feature scaling are performed. In this stage, Value at Risk (VaR) is also calculated to measure the potential financial risk associated with each transaction and added as an additional feature. After preprocessing, the dataset is divided into training and testing datasets to evaluate model performance. Multiple machine learning algorithms, including Logistic Regression, Gradient Boosting, and XGBoost, are then trained and evaluated using the training data. The models are compared based on their accuracy, and the best-performing model is selected and stored along with preprocessing information. In the final stage, users provide transaction details through the user input interface, which are processed by the stored model to generate predictions. The system then outputs the fraud prediction result, classifying the transaction as legitimate or fraudulent, along with a confidence score and alerts to assist users in preventing potential financial fraud.



Fig 5.1: System Architecture

V. IMPLEMENTATION



Fig 6.1: Dataset Uploaded



Fig 6.2: Algorithms trained



Fig 6.3: Machine Learning Model Accuracy Results

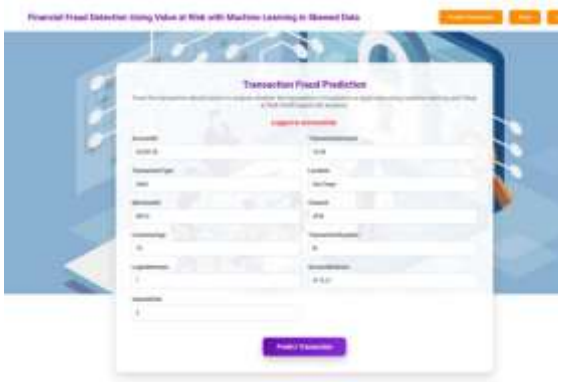


Fig 6.4: Fraud Detected Page.

VI. CONCLUSION

This study presented a machine learning-based framework for financial fraud detection using Value at Risk (VaR) in skewed transaction datasets. The system integrates financial risk analysis with machine learning algorithms to improve the identification of fraudulent transactions. Through data preprocessing techniques such as categorical encoding, feature scaling, and the calculation of VaR, the dataset is enhanced to better represent financial risk patterns associated with transactions.

Multiple machine learning models, including Logistic Regression, Gradient Boosting, and XGBoost, were implemented and evaluated to classify transactions as legitimate or fraudulent. By comparing the performance of these algorithms, the best-performing model was selected to provide accurate fraud predictions. The inclusion of VaR as an additional feature helped capture potential financial risk, thereby improving the model's ability to detect suspicious transactions even in highly skewed datasets where fraudulent cases are relatively rare.

The developed system also provides a web-based interface that allows administrators to manage datasets and train models, while users can input transaction details to obtain fraud predictions along with confidence scores and preventive recommendations. Overall, the proposed approach demonstrates that combining financial risk indicators with machine learning techniques can significantly enhance the effectiveness of fraud detection systems, helping financial institutions reduce financial losses and improve the security of digital financial transactions.

VII. FUTURE SCOPE

The proposed financial fraud detection system can be further enhanced and extended in several ways to improve its performance, scalability, and real-world applicability. Future research can focus on integrating more advanced machine learning and deep learning techniques to improve fraud detection accuracy and adaptability. Algorithms such as Deep Neural Networks, Recurrent Neural Networks (RNN), and Graph Neural Networks (GNN) can be explored to capture complex relationships between transactions and identify hidden fraud patterns more effectively.

Another possible enhancement is the implementation of real-time fraud detection systems that can analyze transactions instantly as they occur. This would allow financial institutions to block suspicious

transactions immediately and prevent financial losses. The system can also be integrated with streaming data processing technologies to handle large volumes of transaction data generated in real-time.

Additionally, future work can incorporate explainable artificial intelligence (XAI) techniques to provide clear explanations for model predictions. This would increase transparency and help financial analysts understand why certain transactions are classified as fraudulent. The system can also be expanded to include additional financial risk metrics beyond Value at Risk (VaR), such as Conditional Value at Risk (CVaR) and other statistical risk measures.

Furthermore, the framework can be deployed as a mobile or cloud-based application, allowing banks and financial institutions to access fraud detection services more efficiently. Integrating the system with blockchain technology and advanced cybersecurity mechanisms could further strengthen the security and reliability of digital financial transactions. Overall, these improvements would make the fraud detection system more intelligent, scalable, and capable of addressing evolving financial fraud challenges.

VIII. REFERENCES

[1] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in *Proc. IEEE Symposium Series on Computational Intelligence*, Orlando, FL, USA, 2015, pp. 159–166. doi:10.1109/SSCI.2015.33.

[2] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "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, 2018. doi:10.1109/TNNLS.2017.2736643.

[3] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *Artificial Intelligence Review*, vol. 34, no. 1, pp. 1–14, 2010. doi:10.1007/s10462-009-9119-y.

[4] J. West and M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Computers & Security*, vol. 57, pp. 47–66, 2016. doi:10.1016/j.cose.2015.09.005.

[5] S. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *Journal of Network and Computer Applications*, vol. 68, pp. 90–113, 2016. doi:10.1016/j.jnca.2016.04.007.

[6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794. doi:10.1145/2939672.2939785.

[7] J. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001. doi:10.1214/aos/1013203451.

[8] D. Hand and W. Henley, "Statistical classification methods in consumer credit scoring: A review," *Journal of the Royal Statistical Society*, vol. 160, no. 3, pp. 523–541, 1997. doi:10.1111/1467-985X.00078.

[9] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi:10.1023/A:1010933404324.

[10] V. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.

[11] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decision Support Systems*, vol. 62, pp. 22–31, 2014. doi:10.1016/j.dss.2014.03.001.

[12] P. Jorion, *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed. New York, NY, USA: McGraw-Hill, 2007.

[13] J. Brownlees and R. Engle, "Volatility, correlation and tails for systemic risk measurement," *Review of Financial Studies*, vol. 30, no. 1, pp. 48–79, 2017. doi:10.1093/rfs/hhw061.

[14] B. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive decision trees," *Expert Systems with Applications*, vol. 42, no. 19, pp. 6609–6619, 2015. doi:10.1016/j.eswa.2015.04.042.

[15] A. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support*

Systems, vol. 50, no. 3, pp. 559–569, 2011. doi:10.1016/j.dss.2010.08.006.