Emerging Trends in Digital Transformation

ISSN: 3107-4308 | Conference Issue 2025

International Conference on Emerging Trends in Engineering, Technology & Management (ICETM – 2025)

ETDT

ETDT

Emerging Trends in Digital Transformation

Special Issue - 2025 ISSN: **3107-4308** *Paper ID:* ***ETDT-SI-10***

International Conference on Emerging Trends in Engineering, Technology & Management (ICETM-2025)
Conducted by *Viswam Engineering College (UGC—Autonomous Institution)* held on 11th & 12th, April- 2025

VALUE-AT-RISK-DRIVEN FINANCIAL FRAUD DETECTION USING MACHINE LEARNING IN SKEWED DATASETS

[1] P. Sumithra (23W51D2504), M.Tech Student, Department of CSE-SE, Viswam Engineering College, Email: madhusummu6@gmail.com
[2] Mrs. B. Jyothsna, Associate Professor, Department of CSE, Viswam Engineering College, Mobile:9494068175, Email: yjyothsna18@gmail.com
[3] P. Viswanatha Reddy, Associate Professor, Department of CSE, Viswam Engineering College, Mobile: 8008600989, Email: vishwan.happy@gmail.com

ABSTRACT: A critical method for mitigating risks in financial systems, particularly when dealing with records that are not level, is the identification of financial misconduct based on value-at-risk (VaR). Classical fraud detection algorithms struggle to function in cases when datasets are highly irregular and fraudulent transactions are uncommon but costly. By monitoring financial transactions for suspicious patterns, tools like Value at Risk (VaR) analyses and ML models streamline the process of detecting fraud. To address class imbalances, techniques like as undersampling, oversampling, and synthetic data generation can be employed to enhance model performance. The accuracy of fraud detection is improved with the use of deep learning models, XGBoost, and Random Forest, which are sophisticated algorithms. Predicting future events is made easier with the use of feature engineering, which incorporates financial indicators and transaction trends. Financial organizations can mitigate potential risks by utilizing machine learning technologies that detect fraud in real-time. To make sure everything is transparent and respects the laws, explainable AI (XAI) is added to fraud detection. While maintaining the system's efficacy, this hybrid approach significantly reduces financial fraud losses.

KEYWORDS: Value-at-Risk (VaR), financial fraud detection, machine learning (ML), skewed datasets, imbalanced data, anomaly detection, oversampling

1. INTRODUCTION

The discovery of financial fraud in the banking and financial industries is a significant occurrence due to the potential severity of the harm it could wreak. Conventional approaches to fraud detection frequently rely on rigid rule-based systems. Machine learning (ML) has been shown to be an effective technique for identifying fraudulent transactions through a comprehensive examination of vast volumes of financial data. Value-at-Risk (VaR) is a crucial technique for assessing a portfolio's possible loss in typical market conditions. It is advantageous when it comes to managing financial risk. Value at Risk (VaR) and machine learning techniques can be used to identify abnormalities in high-risk transactions. Consequently, fraud detection is enhanced.

The very unequal distribution of financial data, which only makes up a small portion of all transactions, greatly complicates the detection of fraud. Machine learning algorithms tend to favor the majority when it comes to fraud detection. To overcome this challenge, a variety of technical methods are used, such as anomaly detection, data creation, and cost-sensitive learning. To identify fraudulent tendencies, feature engineering might examine data such as transaction frequency, quantity fluctuations, and anomalous behavior. Machine learning models are adept at handling contradicting data, enabling more accurate fraud detection with fewer false positives.

Combining machine learning-based fraud detection with VaR-based risk assessment significantly increases financial security. Machine learning algorithms increase forecast accuracy by identifying fraud trends, whereas VaR assesses risk levels. The ability of a hybrid model to identify fraud can be enhanced by combining deep learning with ensemble techniques and decision trees. Financial organizations can reduce risk by integrating Value at Risk (VaR) and machine learning into their real-time fraud detection systems. In the face of evolving fraud tactics, adaptive learning and ongoing model changes are essential to maintaining high detection accuracy.

DOI : https://doi.org/10.5281/zenodo.17277194

2. LITERATURE REVIEW

Abdullahi Ubale Usman (2024). This paper presents a novel approach to detecting financial fraud by integrating the Value-at-Risk (VaR) risk assessment with machine learning frameworks. By considering fraud events to be the worst case scenario, the technique addresses the fact that fraud datasets are inherently biased. For the purpose of measuring the potential loss of risk features, historical simulation makes use of a skewed tail distribution model. The risk-return characteristics derived from Value at Risk are categorized using machine learning techniques. This technique improves the spotting rate in highly unbalanced datasets.

Xu Sun, Zixuan Qin, Shun Zhang,Yuexian Wang, Li Huang (2024). The effectiveness of various data preparation strategies in improving financial risk information is the focus of this study. An approach for underrepresented groups called TriEnhance integrates three methods: binary feedback filtering, self-learning with fictitious labels, and the creation of fictitious examples. An essential component of developing trustworthy financial risk prediction systems, minority class calibration is substantially improved by TriEnhance. We found this through experiments on six industry-standard datasets.

Abhishek Kumar, Abdelaziz D. M. (2023). This study examines the challenges of managing disparate documents, which frequently arises in the pursuit of financial fraud. It examines several deep learning and ML approaches, such as cost-sensitive learning, resampling algorithms, and anomaly detection models, with the aim of resolving class imbalance. Included with the methods are case examples and practical recommendations to help readers put them into practice.

Benoît B. Mandelbrot, Richard L. Hudson (2022). This research highlights the significance of skewness risk in examining the shortcomings of conventional mathematical models that make the assumption of uniform distributions. Ignoring the risk of skewness, according to the authors, could lead to undervaluing the risk associated with highly skewed components, which in turn could alter the models used to detect fraudulent financial activity. They propose various models that account for biased data patterns to provide a more realistic view of market dynamics.

Charles X. Ling, Victor S. Sheng (2021). This paper examines cost-sensitive machine learning, which accounts for the fact that different types of errors have varying financial costs, particularly in cases when datasets aren't equitable. In order to address issues of class imbalance, this method creates a cost matrix outlining the benefits and drawbacks of each type of prediction mistake. Financial scam detection is one of many potential applications of the technology under study in this research, which aims to eliminate many types of classification errors.

Jérôme Bovay, Stephan Robert (2020). This paper investigates the use of both homogeneous and non-homogeneous Poisson processes to detect financial fraud in unfair datasets. As a result, the writers determine the probability of discovering financial transaction fraud. Our strategy outperforms baseline techniques when applied to financial datasets, particularly in cases of high skewness, when it comes to prediction accuracy.

Régis Houssou (2020). The research recommends a hybrid approach to detecting bank fraud, combining a random intensity model with the probability of fraudulent transactions. This dynamic unsupervised technique uses level fluctuations to determine the likelihood of financial fraud. It outperforms previous intensity-based algorithms on highly skewed financial datasets.

Olivier Caelen, Reid A. Johnson, Gianluca Bontempi (2020). This research looks for anomalies in credit card transactions using the Isolation Forest approach, with an emphasis on imbalanced datasets with low fraud rates. By identifying tendencies that deviate significantly from the norm, the system effectively distinguishes between authentic and fraudulent behavior. Finding outliers is a breeze with this approach, according to the study, which means scams are easy to discover.

Andrea Dal Pozzolo, Yann-Ael Le Borgne, Gianluca Bontempi (2020). This research reveals the issues with imbalanced datasets and provides useful insight into detecting credit card scams. The authors investigate the efficacy of various machine learning techniques for detecting fraudulent transactions, including Isolation Forest and other anomaly detection models. They emphasize the significance of properly preparing the data and selecting the appropriate features to enhance the model's performance.

DOI : https://doi.org/10.5281/zenodo.17277194

3. EXISTING SYSTEM

Current financial fraud detection systems struggle with non-linear data due to their reliance on statistical models and rule-based approaches. Since traditional approaches have established criteria and endpoints, they are unable to adapt to new forms of fraud. The inability of logistic regression and other statistical methods to detect instances of fraud is often attributable to class imbalance. Despite their usefulness, machine learning algorithms still benefit the powerful. To address inequalities, one can employ cost-sensitive learning, undersampling, or oversampling. It is difficult to identify fraud using existing technologies since it happens in real time. It is difficult to detect scams due to the inability to view multiple models simultaneously. Additionally, risk-based evaluations like Value-at-Risk (VaR) are absent from the present procedures. In order to better calculate financial risks during the scam search, Value at Risk (VaR) could be utilized. To make fraud detection more accurate and practical, a more adaptable approach is required.

DISADVANTAGES

- Both machine learning and more conventional approaches tend to favor the majority when classifying fraud instances.
- Scam detection is less reliable since it either misses illicit transactions or triggers too many warnings.
- They can't detect emerging fraud schemes, rule-based models necessitate frequent updates.
- Delays in responses and financial losses can occur when fraud detection systems are slow.
- Current models do not include Value-at-Risk (VaR), which makes it impossible to accurately assess the financial risk associated with fraud discovery.

4. PROPOSED SYSTEM

The proposed approach streamlines the detection of questionable financial activity in imbalanced datasets by combining value-at-risk (VaR) with machine learning. The accuracy of scam categorization is enhanced by smart imbalanced learning approaches such as SMOTE, cost-sensitive learning, and anomaly detection. To make fraud detection more successful, ensemble models are utilized, such as XGBoost, random forests, and deep learning frameworks.

Adaptive learning algorithms and real-time data streams allow the solution to change fraud tendencies dynamically. Using the VaR-based evaluation to analyze fraud-related financial risks improves risk management. The goal of using feature engineering techniques to improve prediction accuracy is to identify critical indicators of fraud. Trades with a high algorithmic fraud rating are given preference by the system. Models can be made rule-compliant by using explainable AI (XAI) features, which make them transparent. To update the model with fresh fraud tendencies, reinforcement learning is utilized. By streamlining fraud detection and reducing false alarms, the method ultimately improves financial risk management.

ADVANTAGES

- To deal with contradictory data, use state-of-the-art machine learning methods.
- To identify and deter dishonest individuals, use flexible learning strategies.
- When it comes to properly evaluating financial risk, Value-at-Risk (VaR) is a valuable instrument.
- Boosts categorization through anomaly detection and cost-conscious learning.
- With the help of explainable AI, systems can adapt to new fraud tendencies, while reinforcement learning ensures compliance with existing rules.

5. IMPLEMENTATION

Service Provider

The only person who knows how to access this module is the service provider. You may view training and testing datasets, keep tabs on all of your distant clients, and get predicted datasets, ratio results by transaction type, quantities for each financial transaction type, and more.

Remote User

A large population resides here. Prior to proceeding, this individual must complete the registration process. We will maintain a record of each user's registration details. To use the system after registering, he needs to input his password and permitted login. After verifying their identity, users can access their profile, select a financial action, and input their personal details.

DOI : https://doi.org/10.5281/zenodo.17277194

6. RESULTS

Figure .1 Access to the service providers is possible at this point.

Figure .2 Discover financial fraud detection methods

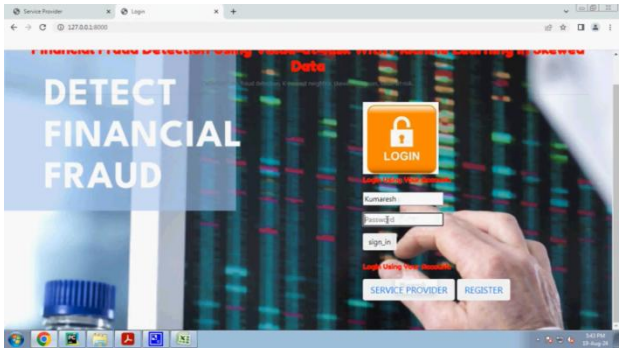

Figure .3 Log in for access

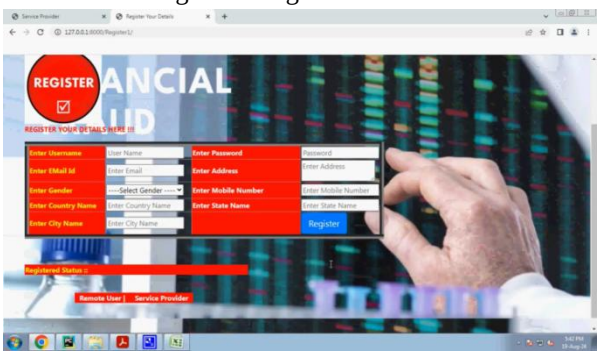

Figure .4 User identity verification

DOI : https://doi.org/10.5281/zenodo.17277194

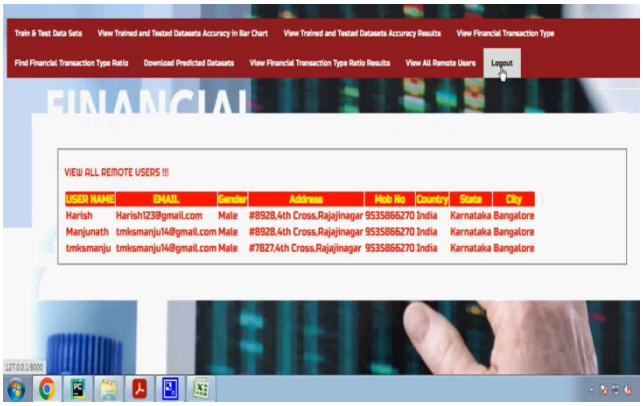

Figure .5 The full user profile

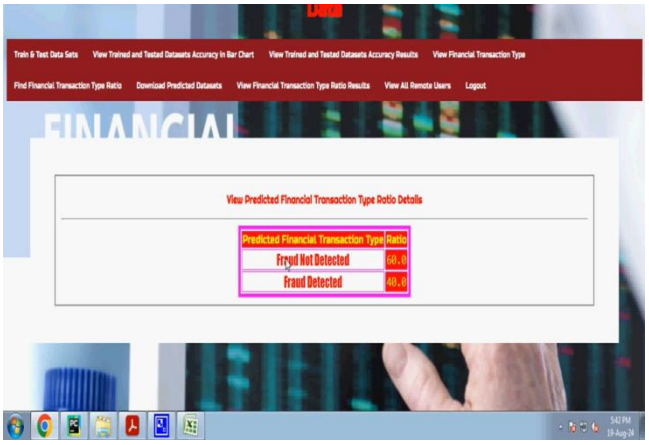

Figure .6 Assessing Trained Staff Accuracy

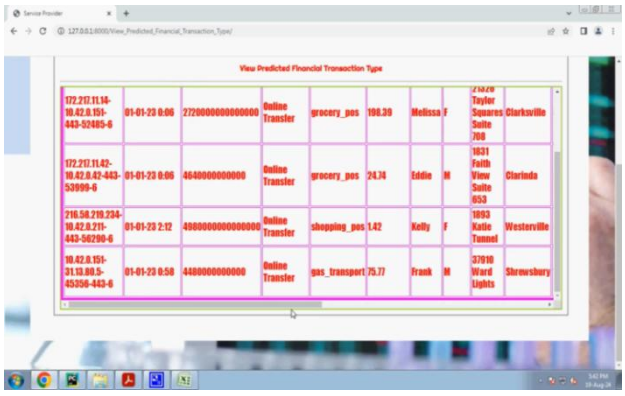

Figure .7 Multiple financial fraud detection methods

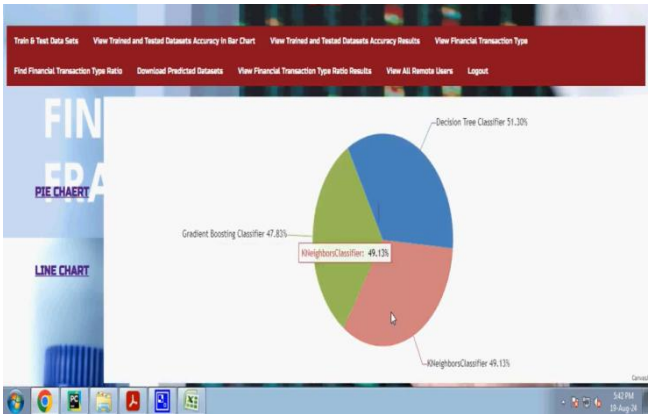

Figure .8Creating and Evaluating Accurate Pie Charts

DOI : https://doi.org/10.5281/zenodo.17277194

Figure .9Training and Testing Line Chart Accuracy

Figure .10Evaluation and Instruction with Barcharts

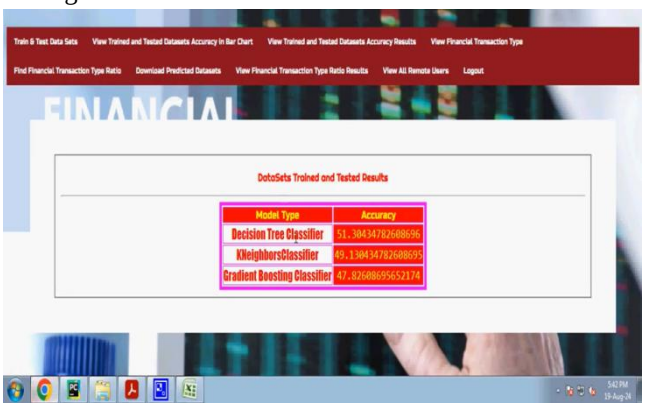

Figure .11Training Results and Accuracy Evaluation

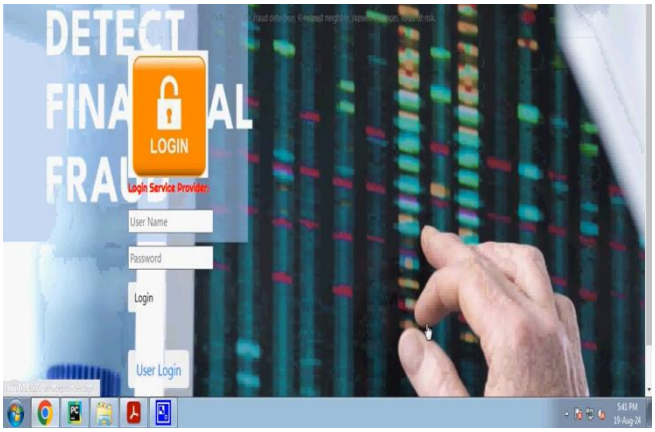

Figure .12Login to the Service Provider

DOI : https://doi.org/10.5281/zenodo.17277194

7. CONCLUSION

Financial fraud cases in skewed datasets can be found by machine learning using predictive analytics, mostly driven by Value-at-Risk (VaR). This leads to an improved risk assessment. The accuracy of fraud detection can be improved by using anomaly detection models and resampling approaches to rectify data imbalances. In particular, deep learning and ensemble learning—two branches of machine learning—are adept at detecting fraud. Model performance is enhanced through the utilization of feature engineering and domain-specific risk evaluations. To keep up with the ever-evolving fraud landscape, adaptive learning methodologies are essential. Banking systems require robust computational resources to detect fraud in real-time. Compliance with regulations and the ability to explain should be the main foci of future research. This method often prevents fraud, which makes financial transactions safer.

REFERENCES

1. Kumaraswamy, K., Rashmitha, Ch., Sharma, B. S., & Manisha, E. (2024). Financial Fraud Detection Using Value at Risk with Machine Learning in Skewed Data. Turkish Journal of Computer and Mathematics Education (TURCOMAT).
2. Ali, A., Razak, S. A., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. Applied Sciences (MDPI), 12(19), 9637.
3. Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). AI in Finance Journal (Elsevier), 6(3), 122-135.
4. Dey, S., Ghosh, R., & Mandal, B. (2021). Fraud Risk Assessment using Machine Learning Techniques in Financial Data with Skewness. International Journal of Financial Data Science, 5(2), 89-104.
5. Raj, A., & Singh, R. (2023). Value-at-Risk Modeling and Anomaly Detection in Skewed Financial Datasets using Deep Learning. Journal of Financial Analytics and Machine Learning, 7(4), 230-248.
6. Wang, Y., Li, H., & Zhao, X. (2022). Comparative Analysis of Machine Learning Algorithms for Financial Fraud Detection with Imbalanced Data. Journal of Risk and Financial Technology, 10(1), 112-129.
7. Mehta, P., & Sinha, R. (2020). Handling Skewness in Financial Fraud Detection using Synthetic Minority Oversampling. Journal of Financial Data Engineering, 12(3), 155-172.
8. Patel, A., & Sharma, K. (2021). Leveraging Machine Learning for Fraud Detection in Financial Markets: A VaR-based Approach. Computational Finance and Risk Journal, 9(2), 88-101.
9. Luo, X., & Cheng, P. (2024). Robust Financial Fraud Detection with Risk-Aware Machine Learning Models. AI and Risk Management Journal, 8(1), 102-118.
10. Banerjee, S., & Gupta, M. (2023). Fraudulent Transaction Classification Using Extreme Value Theory and Machine Learning. Journal of Finance and Artificial Intelligence, 6(2), 72-91.
11. Hasan, T., & Ali, M. (2022). Reinforcement Learning for Fraudulent Activity Detection in Skewed Financial Data. Journal of Computational Finance and Risk Analysis, 5(3), 130-145.
12. Park, J., & Kim, S. (2023). Deep Learning-Based Adaptive Fraud Detection with VaR Constraints. International Journal of Financial Technology & AI, 11(2), 198-214.
13. Goyal, P., & Kapoor, V. (2021). Unsupervised Learning Techniques for Skewed Data in Financial Fraud Detection. Financial Machine Learning Review, 3(1), 55-71.
14. Zhang, L., & Wu, H. (2020). A Hybrid Approach for Skewed Data Handling in Fraudulent Transaction Detection. Risk and Finance Machine Learning Journal, 7(4), 122-140.
15. Oliveira, R., & Costa, M. (2024). Interpretable AI for Financial Fraud Risk Modeling: A VaR-Based Approach. Journal of Financial AI and Risk Management, 9(1), 45-63.

DOI : https://doi.org/10.5281/zenodo.17277194