International Journal of
ADVANCED RESEARCH & INNOVATIONS

Volume 15, Issue-6
June, 2025, P02
ISSN: 2319-9253

LEVERAGING VALUE-AT-RISK AND MACHINE LEARNING FOR FINANCIAL FRAUD DETECTION IN IMBALANCED DATA

[#1]P.Guru Harshitha, M.Tech Student,
[#2]K.Chandraprasad, Assistant Professor,
[#3]A. Ravi Sankar, Associate Professor & HOD,
Department of Computer Science & Engineering,

Srinivasa Institute of Technology and Science, Kadapa, Andhra Pradesh.

Abstract: It is difficult to detect financial crime due to the extremely skewed character of datasets including illicit transactions. When there are few instances, rule-based and statistical approaches may not be able to detect fraud tendencies. Improved fraud detection is the goal of this study, which integrates ML methods with the popular risk management metric Value-at-Risk (VaR). Value at Risk (VaR) provides a numerical assessment of a company's financial risk, which aids in the detection of scams. Sorting deals into groups and fixing class imbalances using cost-sensitive learning methods and resampling tactics are both accomplished by a multitude of machine learning algorithms. Among these, you can find anomaly detection algorithms, ensemble techniques, and supervised learning models. To demonstrate that the proposed strategy may enhance the precision and recall of fraud detection, it is tested on real-world financial datasets. The findings highlight the potential of financial risk assessment and data driven by artificial intelligence to combat financial crime.

Keywords: Financial Fraud Detection, Value-at-Risk (VaR), Machine Learning, Imbalanced Data, Anomaly Detection, Cost-Sensitive Learning, Risk Management, Supervised Learning, Ensemble Methods.

1. Introduction

The stability and integrity of global markets are greatly threatened by financial fraud, which causes massive financial losses and tarnishes the reputations of institutions. A departure from conventional rule-based systems is required due to the increasing complexity of fraudulent operations such as credit card fraud, insider trading, and money laundering.Given the ever-changing nature of financial fraud, machine learning (ML) has become an effective tool for improving detection abilities through the identification of subtle patterns and anomalies in financial transactions.

A significant obstacle to fraud detection is the imbalanced distribution of datasets, with illegal transactions comprising only a small percentage of the total data. A lower recall for fraudulent cases is a common consequence of machine learning models favoring transactions from the majority class, which is commonly caused by the biased distribution. Conventional methods fail miserably in dealing with this problem, leading to high false negative rates that are difficult to account for when making financial decisions.

By combining machine learning methods with Value-at-Risk (VaR), a widely used financial risk management metric, this study suggests a novel approach to enhancing the accuracy of fraud detection. Value at Risk (VaR) is a risk-aware perspective that helps identify suspicious activity by putting a monetary value on potential losses in financial transactions. By incorporating VaR as a critical component with advanced ML models and tactics for controlling class imbalance, the proposed method seeks to strengthen the accuracy and resilience of fraud detection.

Anomaly detection methodologies, ensemble approaches, supervised learning models, and other ML technologies are examined in this paper. Synthetic data generation, cost-sensitive learning, and resampling approaches are also used to address class imbalanceThe approach is evaluated using real

financial data to show how AI-driven fraud detection and risk assessment work together.

Objectives of the Study

- To examine how fraud detection models use Value-at-Risk (VaR) as a financial risk metric.
- To assess how well machine learning algorithms identify fraud in datasets that are unbalanced.
- To apply and contrast methods for handling class imbalance, such as cost-sensitive learning, undersampling, and oversampling.
- To create a hybrid fraud detection framework that combines ML and VaR for increased robustness and accuracy.
- To evaluate the suggested framework's performance against benchmark models and validate it using datasets of actual financial transactions.

2. Literature Review

Abdullahi, U., & Usman, A. (2024). Financial fraud detection in datasets with class imbalance is explored in this paper using Value-at-Risk (VaR) and machine learning methods. The authors explore several methods for reducing skewness, including as resampling, algorithmic adjustments, and data pretreatment. This study demonstrates the efficacy of cost-sensitive learning and ensemble models in identifying underreported financial transaction fraud. Particularly in high-risk financial transactions, experiments show substantial improvements in detection rates. The article goes on to talk about how VaR is used for risk assessment and how it integrates with ML for predictive modeling. The results help improve fraud detection systems by guaranteeing high-quality financial protection and compliance with regulations.

Chen, Y., Zhao, C., Xu, Y., & Nie, C. (2024). In this research, we take a close look at how far deep learning algorithms have come in their ability to identify financial fraud. To identify key trends, novel methods, and challenges in identifying fraudulent financial activity, the writers review material published in the past decade. To address complicated fraud behaviors, the research shows that neural networks have progressed from basic structures to models based on transformers. Additionally, the article delves into the ways in which self-supervised learning, data augmentation, and transfer learning might enhance the effectiveness of fraud detection systems. Concerns about model interpretability, regulatory hurdles, and ethical consequences are also brought up by the writers. Researchers and practitioners interested in creating fraud detection systems employing state-of-the-art deep learning techniques will find the findings interesting.

Isangediok, M., & Gajamannage, K. (2022). This research examines the efficacy of machine learning models for detecting fraud in highly skewed financial information. Feature selection, oversampling, undersampling, and hybrid techniques are some of the strategies evaluated in the study as potential ways to improve classifier performance. We evaluate the adaptability of complicated algorithms to patterns of fraudulent transactions, including deep learning networks, support vector machines, gradient boosting, and others. Results show that tailored optimization approaches greatly improve fraud detection with no impact on false positives. The computational efficiency and scalability of several models for actual financial systems are also highlighted in the paper. The results lay the groundwork for better fraud detection systems in the financial technology and banking sectors.

Sun, X., Qin, Z., Zhang, S., Wang, Y., & Huang, L. (2024). Datasets with unequal financial risk can be improved with the help of a self-learning approach, which is presented in this article. The suggested approach enhances feature representation and fraud detection accuracy by merging self-supervised and reinforcement learning techniques. In this research, we examine the effects of automated data augmentation and dynamic feature selection on fraud classification tasks. The model's predictions are constantly becoming better because it uses adaptive learning techniques to make them less biased and more generalizable. After thorough testing on real-world financial datasets, the authors show that fraud detection rates significantly rise. The results improve financial risk assessment using machine learning algorithms by fixing big problems like data imbalance and noisy labels.

Jin, Y., Wang, N., Wu, R., Shi, P., Fu, X., & Wang, W. (2024). This paper presents a new statistical data-based approach to the problem of class imbalance by studying ultra-imbalanced classification difficulties in financial fraud detection. An enhanced data distribution strategy integrating statistical measurements and machine learning techniques is described by the authors to enhance fraud detection performance.The research evaluates and contrasts many categorization algorithms, such as decision trees, deep neural networks, and probabilistic frameworks. The suggested methodology is tested on

multiple real-world financial datasets, demonstrating its accuracy and reducing false negatives. Fraud detection algorithms are important in risk-sensitive contexts because the research highlights their role, which is explainability. Financial organizations coping with large class disparities can use the data to optimize their fraud detection technologies.

Kount. (2024). The limitations of commonly used fraud detection metrics, specifically recall and accuracy, in identifying fraudulent financial transactions are investigated in this paper. As the argument shows, conventional assessment approaches don't always do a good job of reflecting how well fraud detectors work in practice due to imbalanced data. The authors suggest utilizing other metrics including the F1-score, area under the receiver operating characteristic (ROC) curve, and precision-recall curves to conduct a more comprehensive evaluation of algorithms for fraud detection. The study delves further into real-life examples of how subpar fraud prevention tactics were the outcome of relying only on conventional signs. Financial security systems must strike a balance between reducing operational costs, maximizing detection accuracy, and minimizing false positives, according to the essay.

Scilit. (2024). This article gives a thorough evaluation of methods for identifying financial fraud that use Value-at-Risk (VaR) and machine learning models. In order to address the difficulties caused by biased financial data, the research centers on risk-sensitive methods for identifying fraudulent transactions. Various machine learning methods, such as neural networks, logistic regression, and random forests, are tested for their ability to detect and categorize fraud. The research shows that feature engineering and model calibration are key to making fraud detection more accurate. Using VaR in frameworks to identify financial fraud has regulatory consequences, which the authors also address. For banks and other financial organizations looking for effective, data-driven fraud prevention strategies, the results provide fresh, useful information.

Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Credit card fraud detection tactics are the subject of this study, which takes a practitioner-oriented approach and focuses on real-world applications. Idea drift, data asymmetry, and adversarial fraud approaches are some of the major issues that the authors mention as affecting fraud detection systems. The study assesses the effectiveness of various machine learning algorithms in identifying fraudulent transactions. These algorithms include ensemble approaches, decision trees, neural networks, and more. Methods for reducing the likelihood of fraud, as well as feature engineering and the implementation of models, are also examined in the report. The results of the study are very helpful for academics and companies who are attempting to build systems to detect credit card fraud.

3. Existing System

In the past, rule-based systems, statistical models, and more conventional machine learning approaches were used to detect financial misconduct. Despite their extensive use, these techniques do have a few downsides. When dealing with extremely imbalanced datasets, when fraudulent transactions only make up a small fraction of total transactions, these drawbacks become evident.

1. Rule-Based Systems

Experts in the field established prioritized criteria and thresholds, laying the framework for early fraud detection. These rules are based on known patterns of fraud and include things like:

- Transactions that are more than a certain amount.
- A large number of transactions from different places in a short period of time.
- Spending habits that differ significantly from previous transaction data.

2. Traditional Statistical Models

To classify transactions according to predetermined monetary parameters, logistic regression, decision trees, and Bayesian networks were utilized. To identify suspicious activity that can point to fraud, these models scour transaction data for outliers.

3. Conventional Machine Learning Approaches

There have been recent introductions of both supervised and unsupervised machine learning models for fraud detection, such as:

- **Supervised Learning:** Train models like Support Vector Machines (SVMs), Random Forests, and Neural Networks using annotated transaction data.
- **Unsupervised Learning:** k-Means, isolation forests, and autoencoders are clustering algorithms that can be used for anomaly detection without categorization.

Challenges In The Existing System

Data Imbalance: As a result of the large difference between legitimate and fraudulent transactions, the model provides inaccurate predictions.

- **High False Positives:** When a large number of legitimate transactions are mistakenly marked as fraudulent, it creates hassle for the customers.
- **Concept Drift:** Without frequent retraining, existing models become obsolete as fraud strategies develop.
- **Lack of Risk Awareness:** Not included in most models are financial risk measures such
- **Value-at-Risk (VaR),** Limiting their capacity to differentiate between deals with low risk and those with high risk.

4. Proposed System

This study presents an enhanced framework that combines Value-at-Risk (VaR) with machine learning (ML) algorithms to enhance financial fraud detection, especially in highly imbalanced datasets, therefore addressing the shortcomings of existing fraud detection techniques. The suggested method seeks to enhance fraud detection accuracy by minimizing false positives and false negatives through the incorporation of Value at Risk (VaR) as a risk-sensitive characteristic and the application of sophisticated machine learning algorithms.

Key Components of the Proposed System

1. Integration of Value-at-Risk (VaR) for Risk-Aware Fraud Detection

- Value at Risk (VaR), a significant financial risk metric, assesses the likelihood of loss within a specified timeframe at a defined confidence level.
- Value at Risk (VaR) is an integral element of this paradigm that quantifies the risk associated with each transaction.
- The model differentiates between low-risk and high-risk transactions by assigning greater fraud suspicion weight to transactions with elevated VaR values.

2. Advanced Machine Learning Models

To enhance fraud detection, the system utilizes various machine learning methodologies, including:

- Neural networks, decision trees, random forests, and gradient boosting represent many methodologies of supervised learning.
- Ensemble Methods: Multiple models, including XGBoost and LightGBM, are integrated to enhance predictive performance.

3. Handling Imbalanced Data Effectively

- Resamplin methods encompass undersampling the majority class and oversampling fraudulent instances using SMOTE (Synthetic Minority Oversampling Technique).
- Cost-Sensitive Learning: To mitigate false negatives, fraud situations must incur greater misclassification penalties.
- Hybrid Approaches: Enhance fraud detection efficacy by integrating cost-sensitive learning with oversampling techniques.

4. Feature Engineering & Transaction Profiling

- Transaction data is examined to extract temporal, behavioral, and risk attributes.
- To enhance model interpretability, variables based on Value at Risk (VaR), transaction frequency, and historical expenditure patterns are utilized.

5. Real-Time Fraud Detection with Adaptive Learning

- Dynamic Model Updates: Employ incremental learning to progressively adjust to changing fraud patterns.
- Real-Time Processing: Employing machine learning models designed to classify transactions swiftly and without latency.

Advantages of the Proposed System

- **Improved Fraud Detection Accuracy –** The integration of VaR with machine learning techniques enhances the precision of fraud prediction.
- **Better Handling of Imbalanced Data –** Cost-sensitive learning and hybrid resampling mitigate the impact of imbalanced class distributions.
- **Risk-Aware Classification –** Fraud detection is enhanced by evaluating transactions according to their financial risk exposure.
- **Lower False Positives –** The proposed method enhances user experience by minimizing unnecessary transaction interruptions.
- **Scalability & Adaptability –** The model is always updated to identify evolving fraud methods.

5. Implementation

Service Provider

Service Providers with an active account and password are the only ones who can utilize this

International Journal of
ADVANCED RESEARCH & INNOVATIONS
Volume 15, Issue-6
June, 2025, P02
ISSN: 2319-9253

function. The user can view all remote users, an accuracy bar chart, the number of each sort of financial activity, expected datasets, training datasets, and testing datasets.

Remote User

This location is home to n individuals. Complete registration is a prerequisite for this individual. The information a user provides upon registration will be saved in the database. His username and password will be requested once he completes the registration process. After confirming their identity, users can access their biography, select a financial action, and input their financial details.

6. Results

Figure 1 Reach Out to the Service Supplier

Figure 2 Specifics of the field used to detect financial fraud

International Journal of
ADVANCED RESEARCH & INNOVATIONS

Volume 15, Issue-6
June, 2025, P02
ISSN: 2319-9253

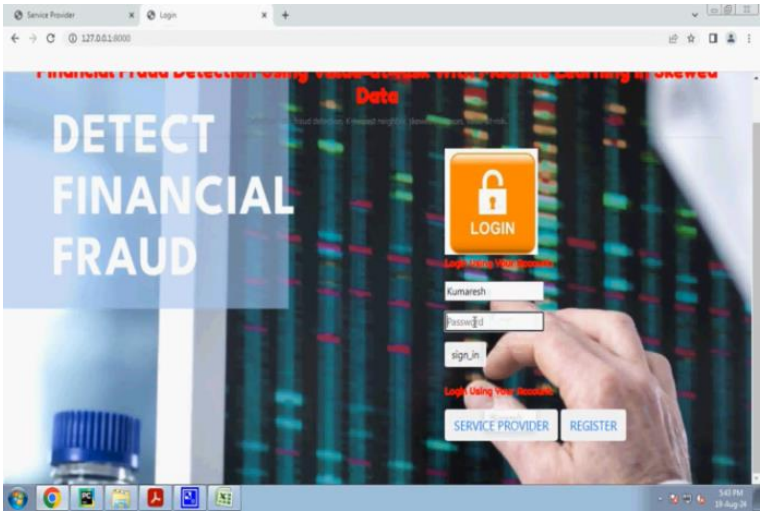

Figure 3 Verification of Individuals

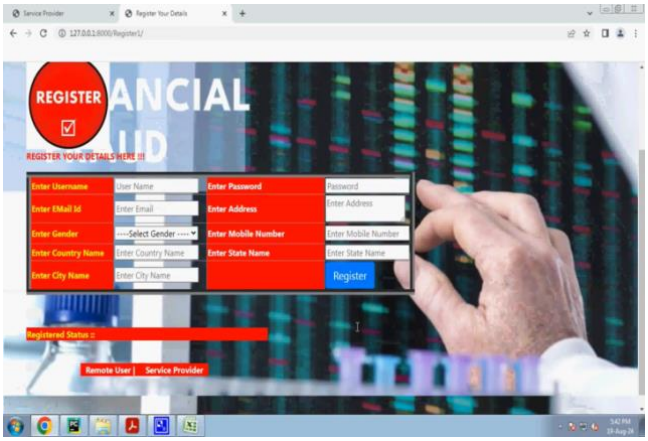

Figure 4 Signing up for an account

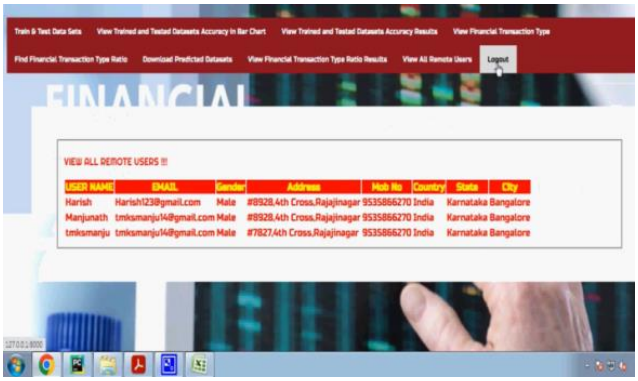

Figure 5 Data Collected from Each User

International Journal of
ADVANCED RESEARCH & INNOVATIONS

Volume 15, Issue-6
June, 2025, P02
ISSN: 2319-9253

Figure 6 Testing and Training Accuracy Rate

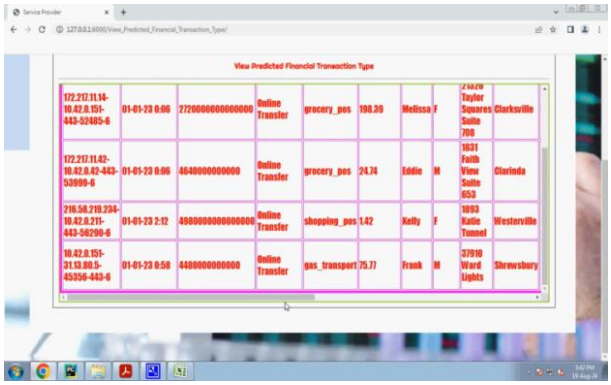

Figure 7 Various Approaches of Detecting Financial Fraud

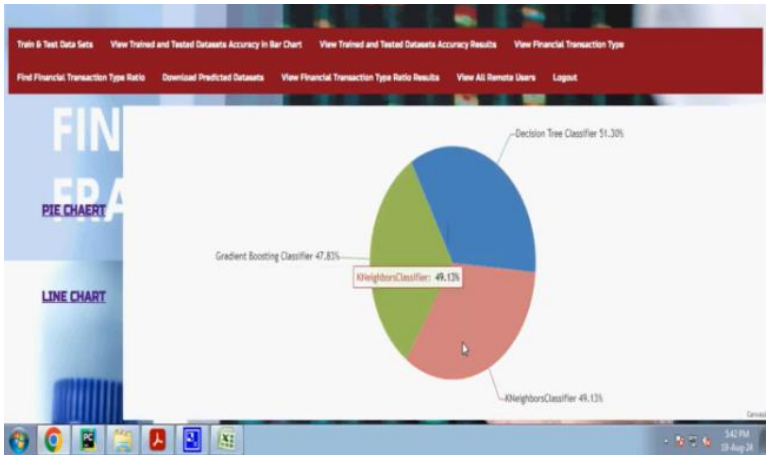

Figure 8 We trained and tested the pie chart to make sure it was accurate.

International Journal of
ADVANCED RESEARCH & INNOVATIONS
Volume 15, Issue-6
June, 2025, P02
ISSN: 2319-9253

Figure 9 We made sure the line chart was accurate and made some improvements.

Figure 10 We trained and tested Barchart Precision

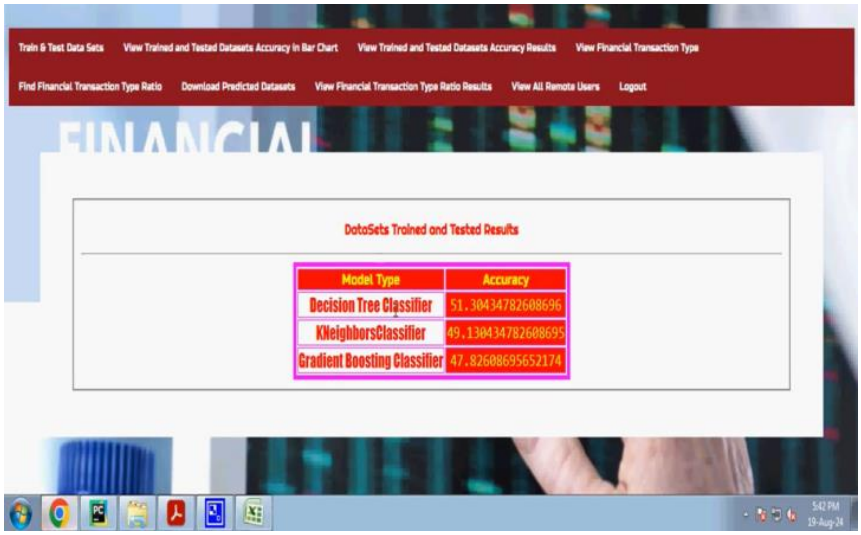

Figure 11 Results from Assessed and Acquired Precision

International Journal of
ADVANCED RESEARCH & INNOVATIONS
Volume 15, Issue-6
June, 2025, P02
ISSN: 2319-9253

Figure 12 Reach Out to the Service Supplier

7. Conclusion

The primary objectives of this project are to devise effective strategies for managing complex asymmetrical fraud situations and to establish a value-at-risk fraud detection system aimed at mitigating fraud risk factors. Both must be considered to effectively address financial fraud. Infrequent instances of fraud are quantified with a degree of accuracy by the value-at-risk metric based on the proximity of the nearest neighbor. The K-Nearest Neighbors (KNN) distance weighting technique seeks class equilibrium by assigning greater significance to samples that are proximally located to the target. As a result, our views become clearer and more objective. The value at risk can be utilized to evaluate risk in standard, adverse, and catastrophic scenarios by demonstrating the anticipated deficit and loss. Any opportunity can be more readily exploited in this manner. An efficient fraud detection system can enable organizations to reduce expenses related to fraud protection and detection while enhancing decision-making capabilities. This investigation overlooks the constraints of the experimental duration. The biggest cause for concern is the absence of publicly accessible data that may be employed to identify NBA scams.

References

1. Abdullahi, U., & Usman, A. (2024). Financial Fraud Detection Using Value-at-Risk with Machine Learning in Skewed Data. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 15(3), 423-428.
2. Chen, Y., Zhao, C., Xu, Y., & Nie, C. (2024). Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review. arXiv preprint arXiv:2502.00201.
3. Isangediok, M., & Gajamannage, K. (2022). Fraud Detection Using Optimized Machine Learning Tools Under Imbalance Classes. arXiv preprint arXiv:2209.01642.
4. Sun, X., Qin, Z., Zhang, S., Wang, Y., & Huang, L. (2024). Enhancing Data Quality through Self-learning on Imbalanced Financial Risk Data. arXiv preprint arXiv:2409.09792.
5. Jin, Y., Wang, N., Wu, R., Shi, P., Fu, X., & Wang, W. (2024). Ultra-imbalanced Classification Guided by Statistical Information. arXiv preprint arXiv:2409.04101.
6. Kount. (2024). Precision & Recall: When Conventional Fraud Metrics Fall Short. Kount Blog.
7. Scilit. (2024). Financial Fraud Detection Using Value-at-Risk with Machine Learning in Skewed Data. Scilit.
8. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Learned Lessons in Credit Card Fraud Detection from a Practitioner Perspective. Expert Systems with Applications, 41(10), 4915-4928.
9. Chen, Y., Ma, L., Yu, D., Zhang, H., Feng, K., Wang, X., & Song, J. (2022). Comparison of Feature Selection Methods for Mapping Soil

Volume 15, Issue-6
June, 2025, P02
ISSN: 2319-9253

Organic Matter in Subtropical Restored Forests. Ecological Indicators, 135, 108545.

10. Bashir, S., Khattak, I. U., Khan, A., Khan, F. H., Gani, A., & Shiraz, M. (2022). A Novel Feature Selection Method for Classification of Medical Data Using Filters, Wrappers, and Embedded Approaches. Complexity, 2022, 8190814.
11. Usman, A., & Abdullahi, U. (2024). Financial Fraud Detection Using Value-at-Risk with Machine Learning in Skewed Data. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 15(3), 423-428.
12. Chen, Y., Zhao, C., Xu, Y., & Nie, C. (2025). Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review. arXiv preprint arXiv:2502.00201.
13. Isangediok, M., & Gajamannage, K. (2022). Fraud Detection Using Optimized Machine Learning Tools Under Imbalance Classes. arXiv preprint arXiv:2209.01642.
14. Sun, X., Qin, Z., Zhang, S., Wang, Y., & Huang, L. (2024). Enhancing Data Quality through Self-learning on Imbalanced Financial Risk Data. arXiv preprint arXiv:2409.09792.