ETDT

Emerging Trends in Digital Transformation

Volume-2, Issue-1, January-March 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I1-13***

FINANCIAL FRAUD DETECTION USING VALUE AT RISK AND MACHINE LEARNING IN IMBALANCED DATA

[#1]P. SWATHI, ***Assistant Professor,***
[#2]B. SHAILAJA, ***Assistant Professor,***
[#3]D. RAHUL, ***Assistant Professor,***
[#4]B. DINESH, ***Assistant Professor,***
Department of Computer Science And Engineering,
TRINITY COLLEGE OF ENGINEERING AND TECHNOLOGY, PEDDAPALLY, TG.

ABSTRACT: The datasets are highly skewed and encompass unlawful activities, which makes it difficult to identify financial crime. When there are few instances of a fraud, rule-based and statistical approaches may fail to detect patterns. The goal of this project is to make it easier to recognize fraud by merging machine learning approaches with the well-known risk management measure Value-at-Risk (VaR). One method for detecting fraud is Value at Risk (VaR), which assigns a numerical value to the financial risk that a company faces. Several machine learning approaches are used to classify deals and correct for class imbalances. Methods for cost-sensitive learning and resampling are among them. Systems that search for anomalous behavior, ensemble approaches, and supervised learning models are among them. The suggested strategy is demonstrated to enhance the accuracy and recall of fraud detection by testing it on real financial datasets.The findings show that AI-driven data and financial risk assessment can be effective tools against financial fraud.

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

1. INTRODUCTION

Financial fraud threatens global markets by destroying companies' reputations and causing massive financial losses. Complexity in credit card fraud, insider trading, and money laundering requires a shift from rule-based systems.Since financial fraud is dynamic, machine learning (ML) can identify subtle patterns and abnormalities in financial transactions to improve detection.

The unequal distribution of datasets makes fraud detection difficult since illicit transactions make up a small part of the dataset. Machine learning algorithms tend to favor the majority class of transactions, reducing their fraud recall. Conventional systems' failure to address this issue can lead to enormous false negative rates that make financial decisions difficult.

This study proposes combining machine learning with Value-at-Risk (VaR), a common financial risk management metric, to improve fraud detection accuracy. Value at Risk (VaR) helps identify suspicious financial transactions by attributing a monetary value to prospective losses. The recommended strategy improves fraud detection by including VaR into advanced ML models and class imbalance management tactics. This paper examines supervised learning models, ensemble approaches, anomaly detection methods, and more.

DOI: https://doi.org/10.5281/zenodo.19016085

Class imbalance is addressed by resampling, synthetic data synthesis, and cost-sensitive learning.Testing on real financial data shows how AI-driven fraud detection and risk assessment work together.

Objectives of the Study

- To investigate the application of Value-at-Risk (VaR) as a financial risk metric in fraud detection algorithms.
- To evaluate how successfully machine learning systems detect fraud in unbalanced datasets.
- To use and compare techniques like cost-sensitive learning, undersampling, and oversampling that address class imbalance.
- To develop a hybrid fraud detection system that boosts accuracy and robustness by combining VaR and ML.
- To validate the proposed system using datasets of real financial transactions and assess its performance against benchmark models.

2. LITERATURE REVIEW

Abdullahi, U., & Usman, A. (2024). This work uses Value-at-Risk (VaR) and machine learning to detect financial fraud in class-imbalanced datasets. The authors try data pretreatment, algorithmic adjustments, and resampling to reduce skewness. We found that ensemble models and cost-sensitive learning can detect underreported financial transaction fraud. Experimental results show considerable detection rate improvements for high-risk financial transactions. The paper then examines VaR's risk assessment applications and how it might be integrated with ML to construct predictive models. Increased financial protection and regulatory compliance result from fraud detection system improvements.

Chen, Y., Zhao, C., Xu, Y., & Nie, C. (2024). We examine deep learning algorithms' financial fraud detection development in this study. Recent decade-long research reveals key trends, creative methods, and challenges in recognizing financial fraud. The study shows how neural networks have evolved from fundamental structures to transformer-based models to tackle complicated fraud. This study examines data augmentation, self-supervised learning, and transfer learning to improve fraud detection systems. The authors discuss ethics, regulation, and model interpretability. Advanced deep learning algorithm fraud detection researchers will benefit from the findings.

Isangediok, M., & Gajamannage, K. (2022). This study examines how machine learning models detect fraud in severely skewed financial data. The study examined feature selection, hybrid approaches, undersampling, and oversampling as classifier performance enhancement strategies. Complex algorithms like gradient boosting, support vector machines, and deep learning networks are examined for fraud adaptation. Data shows custom optimization improves fraud detection with no influence on false positives. The paper examines several financial system architectures for computer efficiency and scalability. The results boost fraud detection systems for banks and financial technology enterprises.

Sun, X., Qin, Z., Zhang, S., Wang, Y., & Huang, L. (2024). This research proposes self-learning to improve unequal financial risk datasets. The proposed method uses self-

DOI: https://doi.org/10.5281/zenodo.19016085

supervised and reinforcement learning to improve feature representation and fraud detection. Our research analyzes how automated data augmentation and dynamic feature selection affect fraud categorization. Adaptive learning makes the model less biased and more generalizable, so it makes better predictions over time. After rigorous testing on real bank data, fraud detection rates improved significantly. Data imbalance and noisy labels are addressed by machine learning algorithms to improve financial risk assessment.

Jin, Y., Wang, N., Wu, R., Shi, P., Fu, X., & Wang, W. (2024). This study examines very unbalanced classifications in financial fraud detection and proposes a statistical data-based solution. To increase fraud detection, the authors propose a new data distribution approach using statistical metrics and machine learning.The research compares decision trees, deep learning neural networks, and probabilistic frameworks for categorization. We evaluate the suggested approach on multiple real-world financial datasets to demonstrate its accuracy and false negative reduction. The study shows fraud detection systems' explainability and significance in risk-sensitive situations. Banks and other financial institutions can detect fraud when class inequalities are large with this data.

Kount. (2024). This study explores the weaknesses of recall and accuracy, two of the most common fraud detection metrics, to identify fraudulent financial transactions. Due to skewed data, present evaluation methodologies may not accurately assess fraud detector performance in real-world situations. The authors recommend utilizing the F1-score, precision-recall curves, and ROC curve area to evaluate fraud detection algorithms. This study explains why standard signal-based fraud prevention fails. Financial security systems must balance detection accuracy, false positives, and operational costs, according to the essay.

Scilit. (2024). Value-at-Risk and machine learning algorithms for financial fraud detection are thoroughly examined in this paper. Research is creating risk-sensitive fraud detection systems to address biased financial data. Neural networks, logistic regression, and random forests are tested for fraud detection and classification. The research reveals that feature engineering and model calibration improve fraud detection. VaR in financial fraud detection frameworks raises regulatory difficulties, which the authors investigate. The results provide fresh, useful data for financial institutions seeking data-driven, efficient fraud prevention.

Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). This practitioner-focused essay examines credit card fraud detection methods as implemented in real life. The authors list concept drift, data asymmetry, and adversarial fraud approaches as major fraud detection system issues. The study compares machine learning systems' fraud detection abilities. This category includes neural networks, decision trees, and ensemble methods. The study discusses feature engineering, model implementation, and fraud prevention. Academics and businesses concerned about identifying credit card fraud will benefit from the study's findings.

3. EXISTING SYSTEM

Conventional machine learning approaches, rule-based systems, and statistical models were once used to detect financial misconduct. These approaches are widely used, however they aren't without their flaws. Due to the small percentage of transactions that are fraudulent,

DOI: https://doi.org/10.5281/zenodo.19016085

these limitations become immediately obvious when dealing with extremely imbalanced datasets.

1. Rule-Based Systems

The development of priority criteria and thresholds by subject-matter experts laid the groundwork for early fraud detection. These rules are based on established patterns of fraud and read as follows:

- Sales that are more than a certain amount.
- Quick processing of a large number of transactions originating from different places.
- Spending habits are very different from historical transaction data.

2. Traditional Statistical Models

Transactions were classified using predefined monetary variables using decision trees, logistic regression, and Bayesian networks. The goal of these models is to detect questionable behavior that could indicate dishonesty by analyzing transaction data for irregularities.

3. Conventional Machine Learning Approaches

There have been recent developments in both supervised and unsupervised machine learning models for fraud detection, including:

- **Supervised Learning:** Make use of annotations in financial transactions to train models like Neural Networks, Random Forests, and Support Vector Machines (SVMs).
- **Unsupervised Learning:** Anomaly detection without classification can be achieved using clustering methods like as autoencoders, isolation forests, and k-Means.

CHALLENGES IN THE EXISTING SYSTEM

1. **Data Imbalance:** Because there is such a large difference between real and fraudulent transactions, the model makes wrong predictions.
2. **High False Positives:** Misclassification of many valid transactions as fraudulent causes frustration for the customers.
3. **Concept Drift:** If existing models are not regularly retrained to account for changing fraud strategies, they will become useless.
4. **Lack of Risk Awareness:** In terms of financial risk measures, most models omit
5. **Value-at-Risk (VaR)**, This limits their ability to discern between high-risk and low-risk trades.

4. PROPOSED SYSTEM

To overcome the limitations of current fraud detection methods, this study suggests a better framework that combines Value-at-Risk (VaR) with ML algorithms. This would improve financial fraud detection overall and in datasets with large biases in particular. By combining modern machine learning techniques with Value at Risk (VaR), a risk-sensitive attribute, the suggested strategy aims to improve fraud detection accuracy by decreasing false positives and negatives.

Key Components of the Proposed System

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

- Valuation at Risk (VaR) is an important financial risk measure that evaluates the probability of loss over a given time frame with a set degree of certainty.

DOI: https://doi.org/10.5281/zenodo.19016085

Volume-2, Issue-1, January-March 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I1-13***

- Important to this paradigm is Value at Risk (VaR), which measures the risk of each transaction.
- To distinguish between low-risk and high-risk transactions, the model gives more weight to transactions with higher VaR values when it comes to suspicions of fraud.

2. Advanced Machine Learning Models

Among the many machine learning techniques used by the system to improve fraud detection are:

- There are a plethora of supervised learning algorithms, including gradient boosting, neural networks, decision trees, and random forests.
- Ensemble Methods: XGBoost and LightGBM are just two of the many models that can be combined to improve prediction performance.
- Two anomaly detection methods, isolation forests and autoencoders, can detect fraudulent transactions even when labelled data is not available.

3. Handling Imbalanced Data Effectively

- To undersample the majority class and oversample fraudulent cases, resampling techniques like SMOTE (Synthetic Minority Oversampling Technique) are used.
- **Cost-sensitive learning**: stronger misclassification penalties should be produced by fraud scenarios to reduce the occurrence of false negatives.
- **Hybrid Approaches**: To improve fraud detection effectiveness, cost-sensitive learning can be combined with oversampling algorithms.

4. Feature Engineering & Transaction Profiling

- Attributes related to time, behavior, and risk are extracted from transaction data.
- Transaction frequency, Value at Risk (VaR), and historical spending trends are included as model variables to improve the model's interpretability.

5. Real-Time Fraud Detection with Adaptive Learning

- Update your model dynamically: use incremental learning to make small but noticeable changes to adapt to changing fraud tendencies.
- The term "real-time processing" describes the use of machine learning models developed for the purpose of transaction classification in a timely manner with minimal delay.

Advantages of the Proposed System

- **Improved Fraud Detection Accuracy** – Combining VaR with machine learning approaches improves the accuracy of fraud prediction.
- **Better Handling of Imbalanced Data** – Uneven class distributions are mitigated via hybrid resampling and cost-sensitive learning.
- **Risk-Aware Classification** – One way to enhance fraud detection is to evaluate transactions according to the level of financial risk they pose.
- **Lower False Positives** – The proposed solution improves the user experience by reducing the occurrence of needless disruptions during transactions.
- **Scalability & Adaptability** – We update the model continually to detect new fraud schemes.

DOI: https://doi.org/10.5281/zenodo.19016085

Volume-2, Issue-1, January-March 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I1-13***

5. IMPLEMENTATION

Service Provider

Only service providers with a current account and password can access this functionality. All remote users, expected datasets, training datasets, testing datasets, and the quantity of each form of financial activity are displayed to the user, along with an accuracy bar chart.

Remote User

There are n people living in this area. Complete registration is necessary for this person. Information regarding a user's registration will be kept in the database. His username and password will be requested once he completes the registration process. Users are able to access their biography, select a financial action, and input their financial details once they have identified themselves.

6. RESULTS

Figure 1 Reach Out to the Service Supplier

Figure 2 Specifics of the field used to detect financial fraud

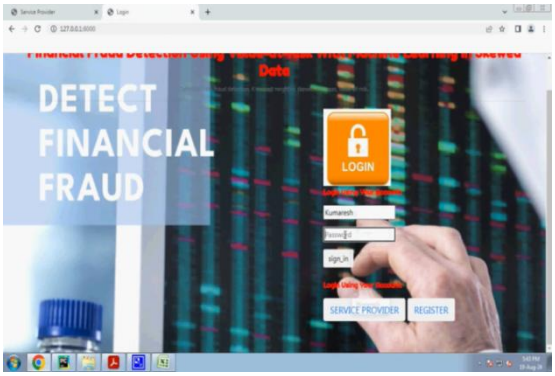

Figure 3 Verification of Individuals

DOI: https://doi.org/10.5281/zenodo.19016085

Volume-2, Issue-1, January-March 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I1-13***

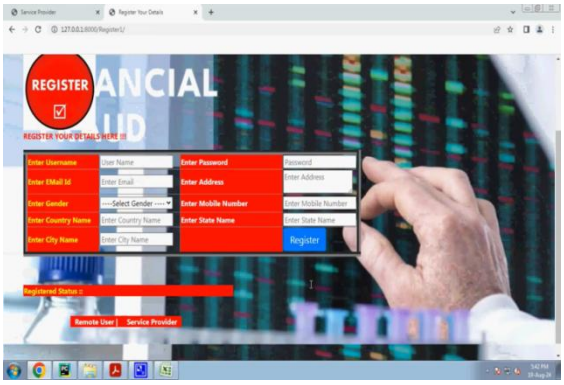

Figure 4 Signing up for an account

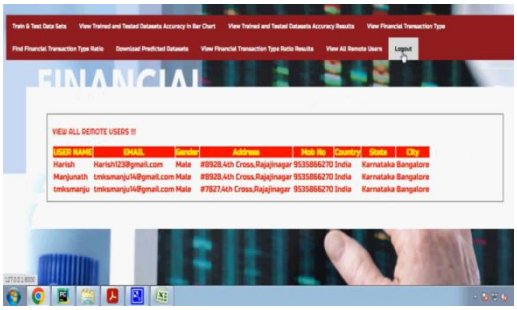

Figure 5 Data Collected from Each User

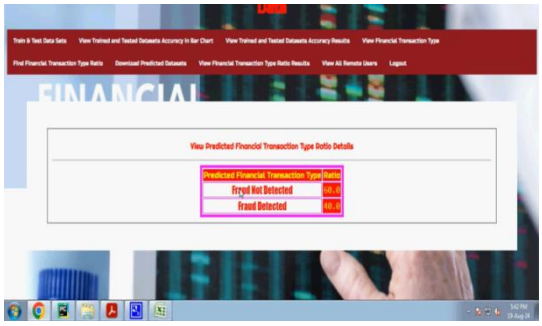

Figure 6 Testing and Training Accuracy Rate

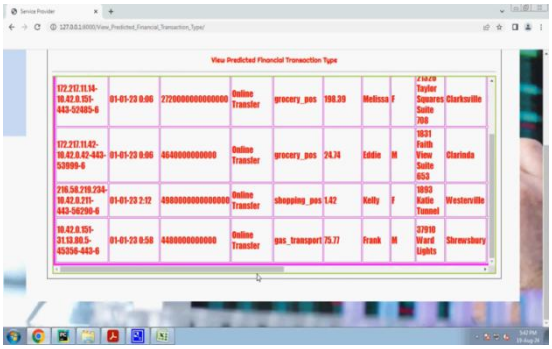

Figure 7 Various Approaches of Detecting Financial Fraud

DOI: https://doi.org/10.5281/zenodo.19016085

Volume-2, Issue-1, January-March 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I1-13***

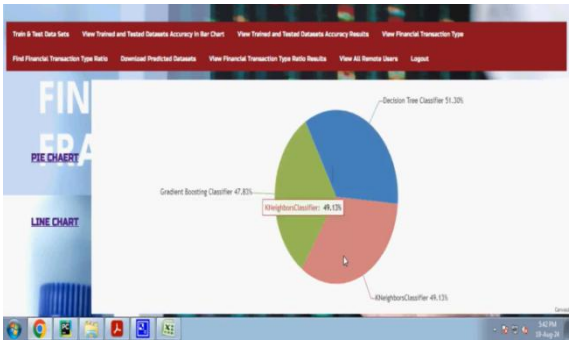

Figure 8 We trained and tested the pie chart to make sure it was accurate.

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

Figure 10 We trained and tested Barchart Precision.

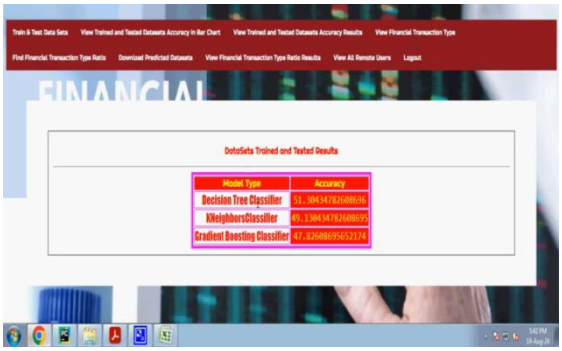

Figure 11 Results from Assessed and Acquired Precision

DOI: https://doi.org/10.5281/zenodo.19016085

Volume-2, Issue-1, January-March 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I1-13***

Figure 12 Reach Out to the Service Supplier

7. CONCLUSION

The primary objectives of this project are to establish a value-at-risk fraud detection system with the purpose of lowering fraud risk factors and to design effective ways for dealing with complex asymmetrical fraud scenarios. Both must be considered in order to fight financial fraud efficiently. Rare instances of fraud can be measured to a certain extent using the value-at-risk metric, which is based on the closeness of the nearest neighbor. The goal of the K-Nearest Neighbors (KNN) distance weighting method is to attain class equilibrium by assigning a higher weight to samples that are closer to the target. This allows us to express our views with greater clarity and objectivity. The value at risk is a useful tool for evaluating risk in favorable, bad, and catastrophic scenarios by depicting the anticipated deficit and loss. Taking advantage of any opportunity becomes much easier with this. An efficient fraud detection system can help businesses save money on fraud protection and detection while also enhancing decision-making abilities. This study disregards the constraints imposed by the duration of the experiment. The primary cause for alarm is the dearth of publicly accessible information that may be utilized 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.

DOI: https://doi.org/10.5281/zenodo.19016085

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 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.