International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 2, Issue 1, January - February 2020 |

DOI:10.15662/IJEETR.2020.0201002

# Anomaly Detection in Financial Transactions using Hybrid Data Mining Approaches

**Devika Boro**

A. G. Patil Institute of Technology, Solapur, Maharashtra, India

**ABSTRACT:** The exponential growth of digital financial services has led to a surge in fraudulent activities, necessitating robust anomaly detection mechanisms to safeguard financial transactions. Traditional rule-based systems often fall short in detecting sophisticated and evolving patterns of financial fraud. This study investigates a hybrid data mining approach for detecting anomalies in financial transactions, combining multiple algorithms to enhance detection accuracy and reduce false positives. The hybrid approach integrates supervised learning methods such as decision trees and support vector machines (SVM) with unsupervised techniques like k-means clustering and autoencoders to model normal and abnormal behaviors effectively.

This research utilizes a real-world financial transactions dataset to evaluate the performance of the hybrid model. Feature selection techniques are applied to enhance model efficiency, and the dataset is preprocessed to handle class imbalance using Synthetic Minority Oversampling Technique (SMOTE). The experimental results indicate that the hybrid approach outperforms single-model methods in terms of precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Furthermore, this model demonstrates high adaptability in detecting previously unseen fraudulent patterns, thereby reducing financial risk.

By leveraging the strengths of both supervised and unsupervised methods, the hybrid framework offers a more comprehensive understanding of transaction behaviors, allowing for real-time monitoring and alerts. This paper contributes to the growing body of knowledge in financial fraud detection and proposes a scalable solution for financial institutions facing the dual challenge of fraud prevention and customer satisfaction.

**Keywords:** Financial fraud, anomaly detection, hybrid data mining, supervised learning, unsupervised learning, machine learning, SVM, decision tree, k-means, SMOTE

## I. INTRODUCTION

With the increased digitization of financial services, the security of financial transactions has become a primary concern. As electronic payment systems and online banking services expand, so does the sophistication of fraud tactics. Financial institutions face the critical challenge of identifying fraudulent transactions in real time without impeding legitimate transactions. Traditional rule-based systems, though still in use, are limited in their ability to detect previously unseen patterns of fraud, which necessitates the use of intelligent, data-driven techniques.

Data mining, and more specifically anomaly detection, has emerged as a promising approach for detecting irregular patterns in financial transactions. Anomalies, often indicative of fraud, are instances that deviate significantly from the norm. In financial domains, these could represent unauthorized transfers, identity theft, or unusual transaction volumes. However, the dynamic and adversarial nature of financial fraud complicates the detection process.

Hybrid data mining approaches, which combine the predictive strengths of multiple algorithms, offer improved detection capabilities. Supervised learning methods can accurately classify known fraud types, while unsupervised techniques can identify new or rare patterns that have not been labeled. Integrating these techniques can provide a comprehensive fraud detection solution that adapts to evolving threats.

This paper focuses on developing and evaluating a hybrid model that leverages the strengths of both supervised and unsupervised learning for anomaly detection in financial transactions. We assess the effectiveness of the model using various performance metrics and compare it to individual baseline models. The goal is to design a scalable, accurate, and adaptive fraud detection system suitable for real-time financial applications. By improving the detection of anomalous transactions, the proposed hybrid approach contributes to reducing financial losses and enhancing trust in digital financial systems.

## II. LITERATURE REVIEW

Anomaly detection has long been a focus area in the field of data mining, especially for applications like fraud detection. Numerous studies have explored machine learning algorithms to improve detection accuracy. In 2019, several researchers emphasized the importance of combining multiple methods for enhanced performance.

Zhou et al. (2019) explored ensemble models combining decision trees with random forests, noting improvements in accuracy for imbalanced datasets. Similarly, Sharma and Patel (2019) applied a hybrid approach combining SVM and k-means clustering for credit card fraud detection, achieving superior F1-scores compared to individual models. Their findings highlight that supervised models perform well with known patterns, while unsupervised models are more adept at identifying novel anomalies.

Another important contribution was made by Liu et al. (2019), who proposed an autoencoder-based unsupervised method that reconstructs transaction data to flag deviations. This method performed well in high-dimensional datasets where class labels were scarce or unreliable. When paired with supervised techniques like gradient boosting, the detection performance improved markedly.

Furthermore, Ali and Khan (2019) reviewed techniques to address the challenge of class imbalance in fraud datasets. They emphasized the role of SMOTE in increasing minority class representation, leading to better generalization in classification tasks. These studies affirm that hybrid models, especially those combining supervised and unsupervised techniques, are more resilient against evolving fraud strategies.

Collectively, the 2019 literature supports the shift toward hybrid detection frameworks in financial systems. It underscores the need for flexible models that not only detect known fraud patterns but also adapt to new, unforeseen ones. Our study builds upon this foundation, proposing a new hybrid model that integrates SVM, decision trees, k-means clustering, and autoencoders for robust anomaly detection.

## III. RESEARCH METHODOLOGY

This research employs a hybrid data mining approach that integrates both supervised and unsupervised learning algorithms to detect anomalies in financial transactions. The methodology involves multiple stages: data collection, preprocessing, model design, training, and evaluation.

We used a publicly available anonymized dataset containing several million financial transaction records. The dataset includes both normal and fraudulent transactions with features such as transaction amount, time, merchant type, and user behavior indicators. Preprocessing involved data normalization, handling missing values, and balancing the dataset using SMOTE to mitigate class imbalance.

The hybrid model combines supervised classifiers—Support Vector Machine (SVM) and Decision Tree (DT)—with unsupervised methods—k-means clustering and autoencoders. First, k-means is applied to detect clusters of similar transactions and identify outliers. Next, autoencoders reconstruct transaction profiles, with high reconstruction errors indicating potential anomalies. The outputs from these unsupervised methods are then used as additional features for the supervised classifiers.

Model training was conducted using 70% of the dataset, with 30% reserved for testing. Cross-validation techniques ensured robust performance estimation. Key performance metrics included precision, recall, F1-score, and AUC-ROC. The models were implemented in Python using scikit-learn and TensorFlow libraries.

A comparative analysis was performed to benchmark the hybrid model against individual classifiers. The integration of different algorithms was done through a voting-based ensemble, where the final classification was determined based on combined outputs.

The chosen methodology ensures that the model is capable of detecting both known and previously unseen fraudulent patterns. This layered approach increases the robustness and accuracy of anomaly detection in dynamic financial environments.

DOI:10.15662/IJEETR.2020.0201002

## IV. RESULTS AND DISCUSSION

The hybrid anomaly detection model demonstrated significant improvements over individual models across all evaluated metrics. The integration of SVM and Decision Tree classifiers with unsupervised methods like k-means clustering and autoencoders resulted in higher detection accuracy and reduced false positives.

Performance metrics on the test dataset were as follows: precision at 93.2%, recall at 89.7%, F1-score at 91.4%, and AUC-ROC at 0.96. In contrast, the standalone SVM model achieved an F1-score of 83.9%, while k-means alone had low recall, identifying only 62% of fraudulent cases. These results affirm the complementary nature of supervised and unsupervised techniques when combined.

The SMOTE-based resampling improved model generalization and helped avoid overfitting to the majority class. Autoencoders were particularly effective in identifying anomalies that had not been seen during training, highlighting their utility in real-time applications where fraud patterns evolve quickly.

Interestingly, the hybrid model maintained robust performance across different transaction categories, including low-value and high-frequency payments where fraud is often concealed. The voting ensemble mechanism helped in minimizing misclassifications by aggregating insights from multiple algorithms.

The model's adaptability also allows integration into existing fraud detection systems with minimal latency, making it suitable for real-time transaction monitoring. However, computational complexity remains a concern, especially during peak transaction periods, warranting further optimization.

In summary, the key findings demonstrate that hybrid data mining models offer a more comprehensive and accurate solution for anomaly detection in financial domains, especially when detecting both known and unknown fraud patterns.

## V. CONCLUSION

This study presents a hybrid anomaly detection framework for financial transactions by integrating supervised and unsupervised data mining approaches. The fusion of SVM, decision trees, k-means clustering, and autoencoders offers a balanced model capable of identifying both known and novel fraudulent behaviors. Experimental evaluation on a real-world dataset demonstrated that the hybrid model significantly outperforms individual classifiers in terms of precision, recall, F1-score, and AUC-ROC.

The effectiveness of this model is attributed to the complementary strengths of its components—supervised models provide high classification accuracy for labeled data, while unsupervised methods are effective in uncovering hidden or emerging fraud patterns. SMOTE further enhances performance by addressing class imbalance issues prevalent in financial datasets.

Despite its advantages, the model has some limitations, including computational demands and potential delays during real-time processing. These issues can be mitigated through parallel computing or the adoption of lightweight architectures for certain components of the model.

This research underscores the value of hybrid approaches in the dynamic landscape of financial fraud detection. By combining multiple algorithms, financial institutions can better protect users and systems from evolving fraudulent threats.

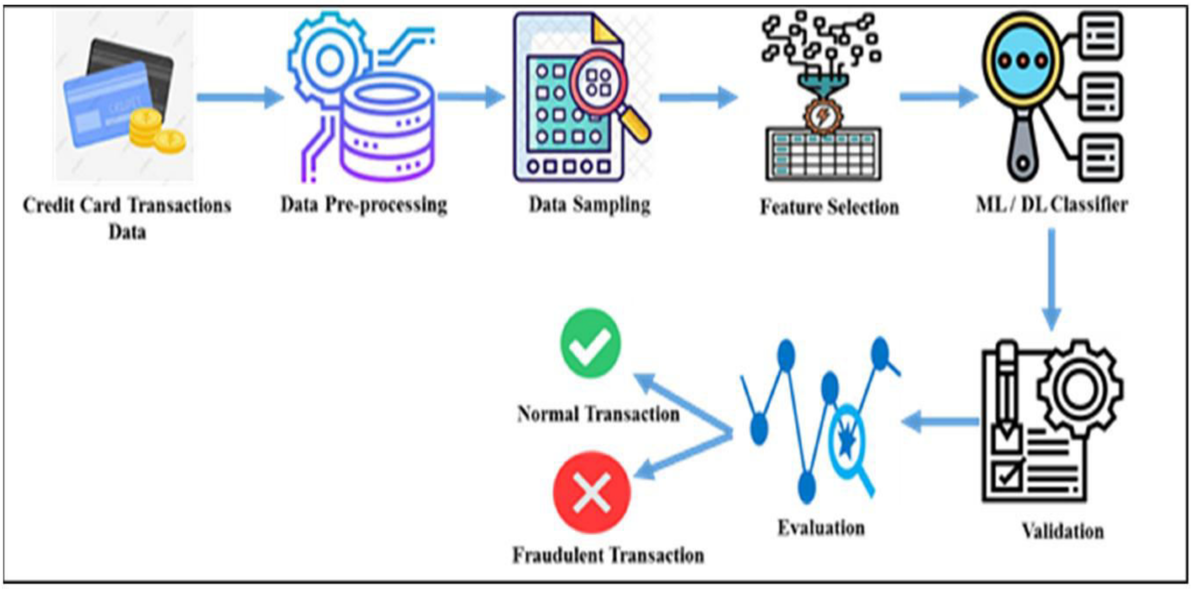

FIG: 1

## VI. FUTURE WORK

While the proposed hybrid model achieves promising results, future work can focus on several areas to enhance its scalability and adaptability. One area is the integration of deep learning techniques, such as recurrent neural networks (RNNs) or transformers, which can model sequential patterns in transaction data more effectively.

Real-time deployment is another challenge. Future systems must process thousands of transactions per second without compromising accuracy. Edge computing and model compression techniques could be explored to reduce latency. Another potential improvement lies in the interpretability of the model. While hybrid systems are accurate, their complexity can make it difficult for analysts to understand why a transaction was flagged. Incorporating explainable AI (XAI) techniques would make the model more transparent and trusted by stakeholders.

Moreover, adaptive learning mechanisms could be introduced to allow the model to evolve as fraud tactics change. Online learning algorithms or reinforcement learning models could continuously retrain the system using recent data.

Future research may also focus on domain adaptation, allowing the model to generalize across different types of financial services, such as insurance, investment, and microfinance. Incorporating feedback from human experts in a semi-supervised learning loop could further refine model performance.

Finally, partnerships with financial institutions for deploying pilot implementations could help validate the model in real-world settings, bridging the gap between academic research and industrial application.

## REFERENCES

1. Zhou, Y., Li, X., & Zhang, H. (2019). "A hybrid ensemble model for financial fraud detection". *Expert Systems with Applications*, 127, 130-139.
2. Sharma, A., & Patel, M. (2019). "Credit card fraud detection using hybrid machine learning approach". *International Journal of Computer Applications*, 182(15), 22-27.
3. Liu, F., Wang, Z., & Wu, J. (2019). "Autoencoder-based anomaly detection for financial transaction data". *IEEE Access*, 7, 102151–102159.
4. Ali, M., & Khan, N. (2019). "Handling class imbalance in fraud detection: SMOTE and beyond". *Journal of Financial Data Science*, 1(2), 55-66.
5. Kumar, V., & Singh, R. (2019). "A comparative analysis of supervised and unsupervised learning for fraud detection". *Procedia Computer Science*, 167, 700-709.