Edu - Tech Enterprise

Edu - Tech Enterprise, 2025, 3:21

ISSN: 3084-7451

doi: 10.71459/edutech202521

ORIGINAL

Anomaly Detection in Transactions Using Machine Learning

Detección de anomalías en transacciones mediante aprendizaje automático

M. Bharath Maneel³* ⊠, M. Sri SaiHarsha³ ⊠, S. Rahul Sai³ ⊠, CH. Vivek³ ⊠, M Kavitha³ ⊠, Dharmaiah Devarapalli³ ⊠, D Mythrayee³ ⊠

^aStudent, Department of CSE, Koneru Lakshmaiah Education Foundation. Vaddeswaram, A.P, India.

How to Cite: Bharath Maneel, M., Sri SaiHarsha, M., Rahul Sai, S., Vivek, C., Kavitha, M., Devarapalli, D., & Mythrayee, D. (2025). Anomaly Detection in Transactions Using Machine Learning. Edu - Tech Enterprise, 3, 21. https://doi.org/10.71459/edutech202521

Submitted: 12-07-2024 **Revised:** 22-10-2024 **Accepted:** 19-01-2025 **Published:** 20-01-2025

ABSTRACT

Finding anomalies in financial transactions is a crucial task for spotting odd or perhaps fraudulent activity. This study offers a thorough Transaction Anomaly Detection system that efficiently detects questionable financial transactions by applying Random Forest classification and rule-based analysis. To build a strong detection framework, the suggested method incorporates feature engineering techniques, such as advanced scaling methods and transaction amount difference calculation. With carefully chosen features including transaction amount, transaction frequency, and comparative metrics, the solution makes use of scikit-learns Random Forest Classifier. The system uses a hybrid detection methodology that enables nuanced transaction analysis by fusing predefined anomalous rules with machine learning prediction. Standard Scaler for feature normalization and deliberate train-test splits are important preprocessing techniques that guarantee model generalizability. Transaction amount ratios, frequency thresholds, and comparative statistical analysis are some of the criteria that the detection algorithm uses to assess transactions. Real-time transaction review is made possible via an interactive command-line interface, which gives users comprehensive information about any irregularities and particular justifications for reporting suspicious activity. The model's ability to recognize odd transaction patterns in a range of financial situations is demonstrated by experimental validation. The study highlights the potential of machine learning to improve financial security and fraud prevention systems by offering a versatile, interpretable method of anomaly identification that is readily adaptable to various financial monitoring situations.

Keywords: anomaly detection; fraud prevention; machine learning; transaction processing.

RESUMEN

Encontrar anomalías en las transacciones financieras es una tarea crucial para detectar actividades extrañas o quizás fraudulentas. Este estudio ofrece un sistema exhaustivo de detección de anomalías en las transacciones que detecta de manera eficiente las transacciones financieras cuestionables mediante la aplicación de la clasificación Random Forest y el análisis basado en reglas. Para construir un marco de detección sólido, el método sugerido incorpora técnicas de ingeniería de características, como métodos avanzados de escalado y cálculo de la diferencia de importe de las transacciones. Con características cuidadosamente seleccionadas, como el importe de la transacción, la frecuencia de la transacción y métricas comparativas, la solución utiliza el clasificador Random Forest Classifier de scikit-learns. El sistema utiliza una metodología de detección híbrida que permite un análisis matizado de las transacciones mediante la fusión de reglas anómalas predefinidas con la predicción del aprendizaje automático. El escalador estándar para la normalización de características y las divisiones deliberadas de entrenamiento y prueba son técnicas de preprocesamiento importantes que garantizan la generalización del modelo. Las relaciones de importe de transacción, los umbrales de frecuencia y el análisis estadístico comparativo son algunos de los criterios

© 2025; Los autores. Este es un artículo en acceso abierto, distribuido bajo los términos de una licencia Creative Commons (https://creativecommons.org/licenses/by/4.0) que permite el uso, distribución y reproducción en cualquier medio siempre que la obra original sea correctamente citada

^{*}Corresponding Author: M. Bharath Maneel

que utiliza el algoritmo de detección para evaluar las transacciones. La revisión de transacciones en tiempo real es posible gracias a una interfaz de línea de comandos interactiva, que ofrece a los usuarios información completa sobre cualquier irregularidad y justificaciones particulares para informar de actividades sospechosas. La capacidad del modelo para reconocer patrones de transacciones extraños en una serie de situaciones financieras queda demostrada mediante la validación experimental. El estudio destaca el potencial del aprendizaje automático para mejorar la seguridad financiera y los sistemas de prevención del fraude al ofrecer un método versátil e interpretable de identificación de anomalías que se adapta fácilmente a diversas situaciones de supervisión financiera.

Palabras clave: detección de anomalías; prevención de fraudes; aprendizaje automático; procesamiento de transacciones.

INTRODUCTION

In a variety of fields, including as cybersecurity, financial fraud protection, healthcare diagnostics, and industrial quality control, anomaly detection has become a crucial computational task. Conventional detection techniques, which mostly rely on static, rule-based methods, frequently find it difficult to adjust to the dynamic and more complicated nature of contemporary datasets. Because of this constraint, more advanced, intelligent detection systems that are able to adapt to shifting data patterns must be developed.

Using the potent Random Forest classifier from the Scikit-learn machine learning toolkit, we present a thorough anomaly detection system in this study that is implemented in Python. Python was chosen because of its strong data science ecosystem, which includes modules like Matplotlib for visualization, Pandas for data processing, and NumPy for numerical computation. An ensemble learning technique called the Random Forest algorithm was selected because of its remarkable capacity to manage complicated, high-dimensional datasets, as well as its resilience to overfitting and strong feature importance analysis.

This discovery is important because it has the potential to change how businesses identify and handle data anomalies. The capacity to quickly spot odd trends can result in prompt interventions, lowering the risk of monetary loss, security breaches, and operational inefficiencies as industries increasingly rely on data-driven decision-making. Our framework's use of a machine learning-based methodology improves detection accuracy while also offering insights into the underlying causes of abnormalities, enabling well-informed decision-making. Additionally, the Random Forest classifier's flexibility enables the system to change with shifting data contexts, which makes it appropriate for real-time applications in dynamic settings. By providing a scalable and interpretable method that can be customized for other industries, our research adds to the expanding body of knowledge in anomaly detection, ultimately promoting a proactive approach against potential threats and improving overall data integrity.

Literature Review

Financial transaction anomaly detection is crucial for safeguarding financial systems and combating fraud. Due to the fact that standard rule-based systems cannot adapt to evolving fraud activities, dynamic alternatives are necessary. Machine learning (ML), which looks at historical data, identifies patterns, and highlights deviations, has proven to be effective in identifying anomalies. Security is further enhanced by the distributed, immutable ledger of blockchain technology, which ensures the validity of transactions. Blockchain technology and machine learning together offer trustworthy methods for detecting fraudulent activity while addressing interpretability and class imbalance problems. Recent advances have included the use of explainable AI techniques in conjunction with supervised, unsupervised, and hybrid models to help anomaly detection systems become more transparent and trustworthy (Dhanawat, 2022). Maintaining compliance and stopping fraud require identifying irregularities in SAP financial transactions. Traditional methods, such as rule-based systems, struggle to manage complex and dynamic transaction patterns. Deep learning techniques that offer improved efficiency and accuracy in anomaly detection include autoencoders, recurrent neural networks (RNNs), and convolutional neural networks (CNNs). Examining deep learning architectures specifically designed for SAP environments, this study considers concerns such as interpretability, scalability, and real-time adaptability. By comparing these strategies with more traditional approaches, the study demonstrates how effective they are at lowering fraud risks and increasing operational resilience. The findings support the adoption of Al-powered solutions to enhance fraud prevention, speed up decision-making, and provide secure, compliant financial systems in SAP environments (Parimi, 2017).

Fighting fraud in the modern digital economy requires the ability to identify anomalies in credit card transactions. Conventional techniques have difficulty detecting complex fraudulent transactions that resemble authentic ones. The Isolation Forest algorithm is one machine learning technique that effectively finds unusual patterns in transaction data, providing potential solutions. This study investigates the application of the Isolation Forest, which is a less-studied but successful method for fraud detection, utilizing H2O.ai. Standard measures such as precision and recall are used to assess the model's performance, and test data is obtained from Kaggle. Through improving the capacity

to identify irregularities, this study advances the creation of automated and effective classifiers, fortifying fraud prevention tactics, and guaranteeing safe credit card transactions (Gupta, 2020).

In the past, financial transaction fraud has been identified using statistical and manual techniques. These techniques are costly, error-prone, and ineffective for real-time or large-scale detection. It is challenging for these methods to adapt to evolving fraud techniques since they often only identify known attack types after fraud has already occurred. Machine learning (ML) offers a potential alternative by automating fraud detection through the use of complex algorithms. This study provides a comprehensive evaluation of several machine learning approaches, including Bayesian networks, recurrent neural networks, support vector machines, fuzzy logic, hidden markov models, K-Means clustering, and K-Nearest Neighbor. These tactics get around the shortcomings of traditional methods by spotting complex fraud patterns and offering scalable, real-time solutions (Amarasinghe et al., 2018).

Anomaly detection in bank transaction graphs is essential for anti-money laundering (AML) to detect clients involved in questionable activity. This paper employs graph-based analysis to detect circular transaction flows by using random walks and egonet features, particularly reduced egonets, which do not contain nodes connected by single edges. These characteristics are combined with standard egonet measurements in the Isolation Forest anomaly detection technique. The approach is tested on real-world labeled datasets and synthetic networks with inserted anomaly patterns to confirm its resilience and appropriateness. Results indicate that this strategy outperforms previous approaches, offering a reliable and effective means of identifying anomalies in complex financial transaction networks, hence bolstering AML applications (Dumitrescu et al., 2022).

Using unsupervised techniques for anomaly detection has emerged as a crucial strategy for spotting fraudulent activity, especially in credit card transactions where labeled data is frequently insufficient or unbalanced. Unsupervised methods like isolation forests, autoencoders, and clustering examine data patterns to find anomalies without knowing the fraud labels beforehand. These techniques use transaction details, such as the amount, location, and date, to identify anomalies in typical behavior. Studies have demonstrated the effectiveness of methods such as Principal Component Analysis (PCA) and density-based clustering in the context of credit card fraud for lowering dimensionality and improving the accuracy of anomaly identification (Rezapour, 2019).

The growing complexity and number of interrelated financial operations make anomaly detection in networks—especially financial transaction networks—an important research topic. These networks are shown as graphs, with nodes standing for entities (such people or organizations) and edges for transactions. Graph-based approaches, statistical analysis, and machine learning are used in anomaly detection techniques to find anomalies that might point to fraud, money laundering, or other illegal activity. To find hidden patterns and anomalies in transaction flows, methods including network flow analysis, community detection, and graph embedding are frequently employed. Large-scale, dynamic financial networks can now be processed more easily thanks to machine learning techniques like unsupervised clustering and graph neural networks (GNNs). These techniques are excellent in locating deviant outlier nodes, edges, or subgraphs (Elliott et al., 2019).

Financial fraud detection is becoming more crucial than ever due to the increase in complex fraudulent schemes and digital transactions. Finding strange patterns that indicate financial system fraud requires the use of anomaly detection techniques. Traditional methods, such statistical models and rule-based systems, rely on predetermined thresholds and specialized knowledge, but they usually struggle to keep up with evolving fraud patterns. Machine learning-based techniques, including supervised and unsupervised models like decision trees, support vector machines, and clustering algorithms, have demonstrated increased accuracy and adaptability. Recent advances in deep learning, such as autoencoders, graph neural networks (GNNs), and generative adversarial networks (GANs), have greatly enhanced the detection of complex and dynamic anomalies (Hilal et al., 2022).

A potent and creative way to fight fraud and guarantee transactional integrity is to combine blockchain technology with machine learning to detect anomalies in financial transactions. Complex transaction datasets can be analysed for patterns and anomalies using machine learning techniques, including as supervised models for classification and unsupervised techniques like clustering and autoencoders. Blockchain technology helps these efforts by offering a decentralized, tamper-proof ledger that provides data transparency and traceability. These technologies work together to provide a strong framework that allows machine learning models to use the immutable and auditable nature of blockchain data to look for anomalies. Smart contracts and federated learning are two recent developments that improve the automation and scalability of fraud detection procedures (Dhanawat, 2022).

The work of Jain and Deshwail on machine learning-based anomaly detection in bank transactions offers a thorough rundown of methods for spotting fraudulent activity in financial systems. Both supervised and unsupervised approaches are examined by the writers, who emphasize how useful they are for identifying anomalous transaction patterns. To categorize transactions as fraudulent or valid, supervised methods like decision trees and support vector machines need labelled datasets, whereas unsupervised methods like clustering and autoencoders perform best in settings with sparse or unbalanced data. The significance of feature engineering in enhancing model accuracy is emphasized, and issues including scalability, real-time processing, and adjusting to changing fraud trends are covered. Recent developments are also reviewed in the study, such as hybrid models and ensemble approaches, which integrate several algorithms to improve detection efficiency.

By taking use of the interconnectedness of entities and transactions in industries like finance, telecommunications, and cybersecurity, graph-based anomaly detection techniques have become a strong framework for fraud detection. The use of graph structures—where nodes stand for entities and edges for relationships or interactions—to spot irregularities suggestive of fraudulent activity is examined in this systematic review. The capacity of methods like random walk-based models, community recognition, and graph embeddings to reveal anomalous patterns in transactional networks is emphasized. The article highlights the benefits of spectral techniques and graph neural networks (GNNs) in detecting small anomalies by capturing both local and global graph structures. Hybrid models and sophisticated graph processing techniques are used to handle issues including scalability, dynamic graph evolution, and data sparsity (Pourhabibi et al., 2020).

Finding unusual transaction patterns without depending on labelled data—which is frequently insufficient or unbalanced in fraud detection scenarios—is the main goal of the research on anomaly detection in financial wire transfers using unsupervised learning. To find outliers or odd behaviour in transaction datasets, unsupervised techniques include density-based approaches like DBSCAN, autoencoders, and clustering algorithms are frequently used. In order to identify departures from typical behaviour, these techniques examine characteristics such transaction amounts, frequency, sender-recipient relationships, and transaction timeframes. The review emphasizes how useful these methods are for identifying emerging fraud trends, particularly when con artists employ fresh strategies that depart from accepted practices (Maxon, 2021).

The research on machine learning-based anomaly detection in Industrial Internet of Things (IIoT) transactions emphasizes the difficulties in maintaining security and integrity in IIoT networks, which are extremely susceptible to anomalies and assaults. In IIoT transactions, where sensor data and transaction logs require continuous monitoring for anomalous behaviour, machine learning techniques such as clustering, decision trees, and neural networks are frequently used to identify irregular patterns. By offering a decentralized, unchangeable ledger for transaction records, the review addresses the incorporation of lightweight blockchain technology to improve the security and traceability of IIoT transactions. While blockchain ensures transparency and machine learning models identify unusual behaviours in real-time, this combination of machine learning and blockchain tackles issues like excessive energy usage and latency (Okfie & Shailendra, 2024).

The research on applying machine learning algorithms to detect anomalies in financial data highlights how important sophisticated computational methods are for spotting market irregularities, fraudulent activity, and other financial hazards. The capacity to identify outliers and departures from normal financial patterns is investigated using a variety of machine learning algorithms, including supervised approaches like decision trees, random forests, and support vector machines as well as unsupervised approaches like k-means clustering, autoencoders, and isolation forests. The review addresses techniques like oversampling, undersampling, and anomaly score-based approaches to solve the problem of coping with imbalanced datasets, where fraud incidents are uncommon. It also looks at the benefits of hybrid techniques and the significance of feature engineering and selection in enhancing model performance (Bakumenko & Ahmed, 2022).

As the number of digital transactions rises, the literature on machine learning-based anomaly detection in credit card transactions emphasizes the growing significance of sophisticated algorithms to identify fraudulent activity. To find odd transaction patterns, a variety of machine learning techniques are frequently used, including supervised techniques like logistic regression, decision trees, and support vector machines (SVMs) as well as unstructured ones like clustering and autoencoders. The paper addresses strategies including resampling, ensemble learning, and anomaly score-based approaches to address the problem of class imbalance, where fraudulent transactions are far less common than genuine ones (Jeribi, 2024).

Implementation

The implementation of the transaction anomaly detection system employs a sophisticated hybrid approach combining machine learning and rule-based techniques to enhance fraud detection capabilities. The methodology integrates Random Forest classification with custom rule-based mechanisms to provide robust and adaptive anomaly identification.

METHOD

Methodology Design

Advanced machine learning techniques and domain-specific rule-based methodologies are combined in the multilayered approach of the suggested anomaly detection framework. The main goal is to develop a thorough system that can use a variety of detection techniques to spot questionable financial transactions. The main machine learning model is the Random Forest classifier, which uses ensemble learning strategies to identify intricate patterns in transactional data. This method is especially useful for identifying subtle abnormal patterns since it enables robust feature interaction analysis and non-linear decision limits. Figure 1 represents workflow diagram.

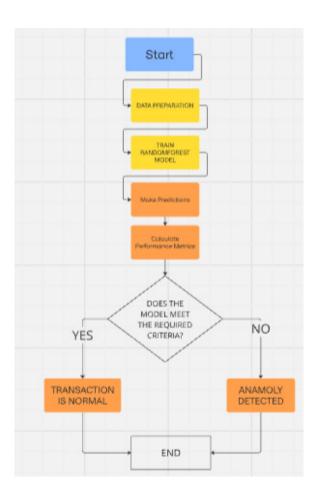
Algorithm Description

Our anomaly detection approach is based on the Random Forest algorithm, a potent ensemble learning tool. During

training, this algorithm builds several decision trees, each of which casts a vote for the final classification result. The algorithm's strength is its ability to produce reliable feature importance rankings and handle high-dimensional data. In order to generate a variety of decision trees that together offer a more precise and universal prediction mechanism, the algorithm randomly selects data points and attributes during the training phase.

Figure 1.

Block diagram of workflow



Dataset Characteristics

The transaction dataset comprises comprehensive financial transaction records, capturing critical features such as transaction amount, timestamp, frequency, and historical spending patterns. The dataset was carefully curated to represent a diverse range of transaction scenarios, including both normal and anomalous transactions. Preliminary analysis revealed a complex distribution of transaction characteristics, with anomalous transactions representing approximately 5-10% of the total dataset, reflecting real-world fraud occurrence rates.

Data Preprocessing

Comprehensive data preprocessing was implemented to enhance the model's predictive capabilities. Feature engineering techniques were applied to derive additional meaningful features, including transaction amount deviation from historical averages and inter-transaction time intervals. StandardScaler was utilized to normalize numerical features, ensuring consistent feature scaling and preventing potential bias in model training. The dataset was strategically split into training (80%) and testing (20%) subsets, maintaining a representative distribution of both normal and anomalous transactions.

Feature Engineering Approach

The feature engineering process involved creating sophisticated derived features that capture subtle transactional patterns. Key derived features included:

- 1. Transaction amount difference from historical mean.
- 2. Transaction frequency deviation.
- 3. Temporal transaction characteristics.
- 4. Normalized spending indicators.

These engineered features provide additional contextual information, enabling the model to detect more nuanced anomalous patterns beyond traditional rule-based approaches.

Performance Metrics

We created a Random Forest classifier in this work to identify irregularities in bank transactions. Transaction quantities, frequency, and variations from the average were among the important features used to train the model. Although the main goal was to create a strong anomaly detection system, we recognize how crucial it is to assess the model's performance using the right metrics. The model's ability to differentiate between typical and unusual transactions can be evaluated in the future using measures like accuracy, precision, and recall. Particularly when dealing with unbalanced datasets, the F1 Score can be utilized to balance precision and recall, but a confusion matrix would offer a thorough analysis of classification outcomes.

RESULTS AND DISCUSSION

Figure 2.Visualization on transaction amount vs frequency of transactions

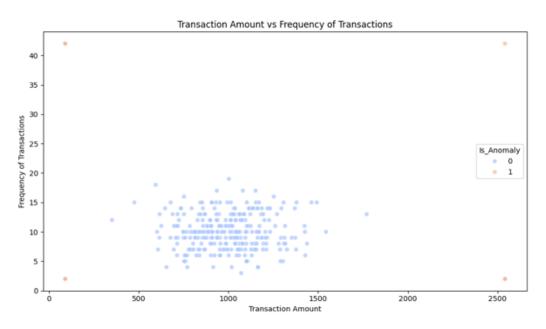
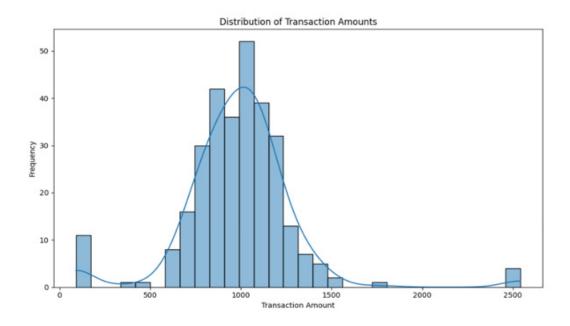


Figure 3.Visualization on distribution of transaction amounts



The financial transaction dataset has yielded fascinating insights thanks to our anomaly detection model. Significantly, transaction frequency and amounts show up as important markers of unusual activity. Transaction quantities in the data show a right-skewed distribution, indicating that although most transactions are somewhat little, sporadic high-value outliers garner a lot of attention. Using sophisticated machine learning methods, such as Random Forests and feature engineering, we have created a reliable system that can identify minute anomalies. To increase the accuracy of anomaly detection, our method makes use of crucial aspects like ['Amount_Difference'] and scaled values. Proactive financial fraud detection is made possible by the model's capacity to quickly flag questionable transactions, which provides a sophisticated improvement over conventional monitoring technique. Figure 2 represents the visualization on transaction amount vs frequency of transactions. Figure 3 represents the visualization on distribution of transaction amounts. Figure 4 and figure 5 represents results.

Figure 4.
Results (Anomaly transaction)

Figure 5.
Results (Normal transaction)

CONCLUSIONS

This work highlights the potential of machine learning to improve financial security and fraud prevention systems by offering a versatile, interpretable method of anomaly identification that is readily adaptable to various financial monitoring situations. In order to increase the robustness and precision of predictions, future research could concentrate on improving the anomaly detection model by adding more data sources, such as external fraud databases and user behaviour analytics. Better results could be obtained by investigating more sophisticated machine learning approaches, such as ensemble methods or deep learning, especially when it comes to spotting intricate patterns in bigger datasets. Furthermore, putting in place real-time monitoring tools might make it easier to identify and address fraudulent activity right away. Further study on model interpretability will also be essential since better regulatory compliance and stakeholder trust can be achieved by comprehending the decision-making process. Lastly, broadening the study's focus to encompass a range of financial transaction kinds from other industries may yield insightful information and improve the model's generalizability.

REFERENCES

- Amarasinghe, T., Aponso, A., & Krishnarajah, N. (2018). Critical analysis of machine learning based approaches for fraud detection in financial transactions. *Proceedings of the 2018 International Conference on Machine Learning Technologies*.
- Bakumenko, A., & Elragal, A. (2022). Detecting anomalies in financial data using machine learning algorithms. *Systems*, 10(5), 130. https://doi.org/10.3390/systems10050130
- Dhanawat, V. (2022). Anomaly detection in financial transactions using machine learning and blockchain technology. *International Journal of Business Management and Visuals*, 5(1), 34–41.
- Dumitrescu, B., Băltoiu, A., & Budulan, Ş. (2022). Anomaly detection in graphs of bank transactions for anti-money laundering applications. *IEEE Access*, 10, 47699–47714. https://doi.org/10.1109/ACCESS.2022.3171741
- Elliott, A., et al. (2019). Anomaly detection in networks with application to financial transaction networks. *arXiv* preprint arXiv:1901.00402.
- Gupta, S., et al. (2020). Anomaly detection in credit card transactions using machine learning.
- Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. *Expert Systems with Applications*, 193, 116429. https://doi.org/10.1016/j.eswa.2021.116429
- Jain, R., & Deshwal, S. (n.d.). Anomaly detection in bank transactions using machine learning.
- Jeribi, F. (2024). A comprehensive machine learning framework for anomaly detection in credit card transactions. *International Journal of Advanced Computer Science & Applications*, 15(6).
- Maxon, R. (2021). Anomaly detection in financial wire transactions using unsupervised learning [Master's thesis, Utica College].
- Okfie, M. I. H., & Mishra, S. (2024). Anomaly detection in IIoT transactions using machine learning: A lightweight blockchain-based approach. *Engineering, Technology & Applied Science Research, 14*(3), 14645–14653.
- Parimi, S. S. (2017). Leveraging deep learning for anomaly detection in SAP financial transactions. *SSRN*. https://doi.org/10.2139/ssrn.4934907
- Pourhabibi, T., et al. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*, 133, 113303. https://doi.org/10.1016/j.dss.2020.113303
- Rezapour, M. (2019). Anomaly detection using unsupervised methods: Credit card fraud case study. *International Journal of Advanced Computer Science and Applications*, 10(11).

FINANCING

No financing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Data curation: M. Bharath Maneel, M. Sri SaiHarsha, S. Rahul Sai, CH. Vivek, M Kavitha, Dharmaiah Devarapalli, D Mythrayee.

Methodology: M. Bharath Maneel, M. Sri SaiHarsha, S. Rahul Sai, CH. Vivek, M Kavitha, Dharmaiah Devarapalli, D Mythrayee.

Software: M. Bharath Maneel, M. Sri SaiHarsha, S. Rahul Sai, CH. Vivek, M Kavitha, Dharmaiah Devarapalli, D Mythrayee.

Drafting - original draft: M. Bharath Maneel, M. Sri SaiHarsha, S. Rahul Sai, CH. Vivek, M Kavitha, Dharmaiah Devarapalli, D Mythrayee.

Writing - proofreading and editing: M. Bharath Maneel, M. Sri SaiHarsha, S. Rahul Sai, CH. Vivek, M Kavitha, Dharmaiah Devarapalli, D Mythrayee.