SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Financial Fraud Detection Using Isolation Forest And DBSCAN Clustering Techniques

K. Anitha[1*], R. Vinoth[2], Vinay Kumar Deolia[3], Dr.R. Siva Thivya[4], Dr.S.G. Balakrishnan[5], Dr. Sasi Bhushan[6]

[1*]Associate Professor, Department of Management Studies, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: anithak@maher.ac.in, https://orcid.org/0000-0002-1940-2101
[2]Asst Professor, Department of Management Studies, New Prince Shri Bhavani College of Engineering and Technology, Chennai, Tamil Nadu, India. E-mail: vinoth190495@gmail.com, https://orcid.org/0009-0008-8729-3489
[3]Department of Electronics & Communications Engineering, GLA University, Mathura, Uttar Pradesh.
E-mail: vinaykumar.deolia@gla.ac.in, https://orcid.org/0000-0003-0975-2121
[4]Assistant Professor, KPR College of Arts Science and Research Coimbatore, Tamil Nadu, India.
E-mail: shivasweety843@gmail.com, https://orcid.org/0000-0002-0528-3789
[5]Professor, Computer Science and Engineering, Mahendra Engineering College, Namakkal, Tamil Nadu, India.
E-mail: balakrishnansg@mahendra.info, https://orcid.org/0000-0001-5737-6103
[6]Department of MBA, Ramachandra College of Engineering, Eluru, India E-mail: sasibhushan@rcee.ac.in
*Corresponding author: Email: anithak@maher.ac.in

Abstract

Financial fraud is an urgent problem in modern banks and fintech companies that leads not only to significant financial losses but also undermines customer trust in these companies. Rule-based methods and supervised learning (SL) models have some limitations in recognizing new cases and infrequent incidents. Hence, the research aimed to develop a hybrid architecture based on Isolation Forest (IF) and DBSCAN algorithms to improve the efficiency of fraud detection while providing the interpretability of results. The model was tested using the Kaggle Credit Card Fraud dataset that included 284,807 entries with highly unbalanced classes (fraudulent cases = 492). The preprocessing step included normalization of the Time and Amount variables. Initially, IF identified outliers. The next step was to apply DBSCAN clustering using those anomalies. Different measures were used for performance analysis of the proposed approach. The values of 99.88% for accuracy, precision 0.93, recall 0.87, F1 score 0.90, and ROC-AUC score 0.96 were achieved using the hybrid model. From the cluster evaluation results, the former clusters are clearly separated with Silhouette score being 0.61 and the DBI value of 0.45. It is evident that the hybrid approach has produced understandable results which can be used for fraud detection and grouping. Through comparative study, it is clear that the hybrid approach outperforms the IF. It shows that combination of IF and DBSCAN will be effective for fraud detection due to interpretability.

Keywords: Credit Card Fraud, Isolation Forest, DBSCAN, Anomaly Detection, Clustering, Hybrid Framework, Financial Risk Management.

1. Introduction

Fraud is an ongoing issue that causes massive financial damage globally. While it threatens the financial stability of the entities concerned, it also shakes the trust of its customers in its banking and fintech facilities [6]. With the increasing use of digital payments and mobile banking, among other internet-based financial products, the

sophistication of the fraud attempts also increases [7]. It has become ever more important, therefore, that effective solutions must be developed to counter such a growing problem.

Identification of cases of financial fraud is important as far as ensuring minimal financial loss and risks as well as ensuring integrity of financial processes is concerned [8] [10]. This process makes it possible for institutions to act on time without the risk of the situation becoming worse and helps to comply with relevant regulations. The early identification process also minimizes the risk of reputation and operational costs due to investigations.

Techniques associated with unsupervised machine learning (ML) provide an efficient method of dealing with fraud when data available is limited, inconsistent, or ever-changing. Algorithms such as IF and DBSCAN clustering can identify fraud and find patterns within a set of data [9] [16]; thus, offering solutions which are much more dynamic and flexible than those provided by traditional rule-based systems [1] [4]. Unsupervised ML techniques also allow for continuous learning and adaptation to emerging trends and patterns within data sets, making them well-suited for environments where fraud is continually evolving.

The main aim of this study is to create a hybrid model using IF for anomalies and DBSCAN for clustering in order to detect fraudulent transactions. In addition to improving detection efficiency and revealing hidden patterns in the data, the creation of this model should make it possible to use its outcomes for decision-making regarding transaction validation. Another important advantage of the suggested hybrid model is its interpretability as related to some other techniques.

The paper organization is as follows: Section II offers a literature review regarding anomaly detection, clustering algorithms, the IF algorithm, the DBSCAN algorithm, and the hybrid approach. Section III describes the data set, data pre-processing, proposed hybrid method, and performance evaluation criteria. Section IV provides experimental methodology and results analysis. Section V highlights the conclusions, practical implications, and limitations of the study. Section VI completes the paper with contributions and directions of future research.

2. Literature Survey

Many researchers have investigated financial fraud detection using different ML techniques that can be categorized into three main categories: supervised, unsupervised, and hybrid methods [11] [12]. SL involves the use of labeled datasets to classify legitimate from fraudulent transactions [14] [15]. Although supervised methods can deliver high accuracy rates, these methods depend on the availability of labeled data, which are rare in practice. Unsupervised learning detects anomalies without any previous knowledge, thus useful in detecting dynamic fraud patterns [5] [13]. However, unsupervised models tend to produce more false alarms than supervised models. Hybrid approaches integrate the advantages of both supervised and unsupervised learning by exploiting labeled data when possible while dealing with unseen anomalies [17].

IFs have gained popularity among many researchers as a means of detecting anomalies in financial data without any supervision. Through recursive partitioning, it can isolate the suspicious transactions. Clustering methods have also been useful for grouping similar transactions together while isolating suspicious transactions as anomalies [18]. Research on fraud detection using IF and Clustering has proved successful through their use in different financial institutions, including banks, credit card services, and e-payment systems [2][3].

Several challenges have been faced with these approaches to fraud detection [19] [20]. Rare fraud patterns are difficult to detect, especially as they keep changing [21] [22]. Most of these methods lack the scalability necessary to work efficiently in an environment with large transaction volumes. Black-box models used in many of these approaches limit their usefulness due to low interpretability.

The problems mentioned above are the driving forces behind the requirement of integrated solutions, which will integrate anomaly detection along with clustering techniques to make the algorithms more precise and interpretable. The grouping of the IF and DBSCAN algorithms can solve these issues since the IF algorithm will detect the suspicious transactions, and the DBSCAN algorithm will cluster them.

3. Methodology

3.1 Description of Dataset

The dataset chosen for this experiment is the freely available https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud from Kaggle, consisting of 284,807 credit transactions carried out in two days by customers from Europe. Each credit transaction consists of 30 numeric attributes, where 28 attributes correspond to the anonymized principal component attributes (V1-V28) obtained through the process of Principal Component Analysis. Apart from those attributes, each transaction includes Time and Amount (value of transaction). The target

variable named Class categorizes each transaction either as fraudulent (1) or genuine (0), with only 492 cases marked as fraudulent (0.172%). Standardization and normalization of the Time and Amount attributes were performed during data preprocessing, while no transformation was necessary for the remaining attributes.

3.2 Anomaly Detection using IF

IF algorithm was implemented to achieve primary anomaly detection in an unsupervised learning approach. This algorithm helps to detect possible outliers through recursive partitioning of the dataset based on the shortest path lengths required to isolate transactions. Those transactions which require fewer splits in the tree to isolate from other observations in the dataset are considered outliers. The number of estimators and contamination were optimized via cross-validation to achieve sensitivity. The IF method is ideal for analyzing financial data because it is efficient at identifying outliers in large datasets without labels.

IF detects anomalies by isolating observations using recursive random partitioning. The anomaly score for a transaction $x_i$is computed as equation (1):

$$s(x_i, n) = 2^{-\frac{E(h(x_i))}{c(n)}} \quad (1)$$

Where:

- $E(h(x_i))$represents the average path length of $x_i$across all trees,
- n represents the number of samples in the dataset,
- $c(n) = 2H(n-1) - \frac{2(n-1)}{n}$ represents average path length of unsuccessful searches in a Binary Search Tree,
- $H(i) = \sum_{k=1}^{i} \frac{1}{k}$ represents the harmonic number.

An anomaly is identified if $s(x_i, n) \approx 1$; normal transactions have $s(x_i, n) < 0.5$.

3.3 DBSCAN Clustering Method

The DBSCAN algorithm is used to group together similar transactions and identify clusters of potentially fraudulent behaviors. The DBSCAN algorithm requires two main inputs: the parameter ε, which is the radius of the neighborhood, and the parameter minPts, which represents the minimum number of points necessary to constitute a cluster. Any transaction that not belong to any dense region will be identified as noise. The parameters were chosen using k-distance graphs in order to maximize cluster separation while still being able to understand the clustering structure.

DBSCAN groups transactions based on density. A point $x_i$is a core point of:

$$| N_\epsilon(x_i) | \geq \text{minPts} \quad (2)$$

Where in equation (2), $N_\epsilon(x_i) = \{ x_j \mid d(x_i, x_j) \leq \epsilon \} represents$ the set of points within a radius $\epsilon$of x_i, and $d(\cdot)$is typically Euclidean distance.

Points that are not core points but are within $\epsilon$of a core point are border points, and points that are neither are considered **noise** (potential anomalies).

3.4 Proposed Hybrid Workflow

This hybrid framework involves the implementation of IF together with DBSCAN to enhance the performance of the fraud detection algorithm. Here, IF initially detects the anomalies from the entire dataset. The detected anomalies are then clustered using DBSCAN to unveil any hidden patterns and groupings of fraud transactions. This two-step technique ensures more accurate fraud detection and can be used to detect new fraud patterns and provide valuable information to financial risk analysts.

Figure 1: Hybrid IF–DBSCAN Workflow for Credit Card Fraud Detection

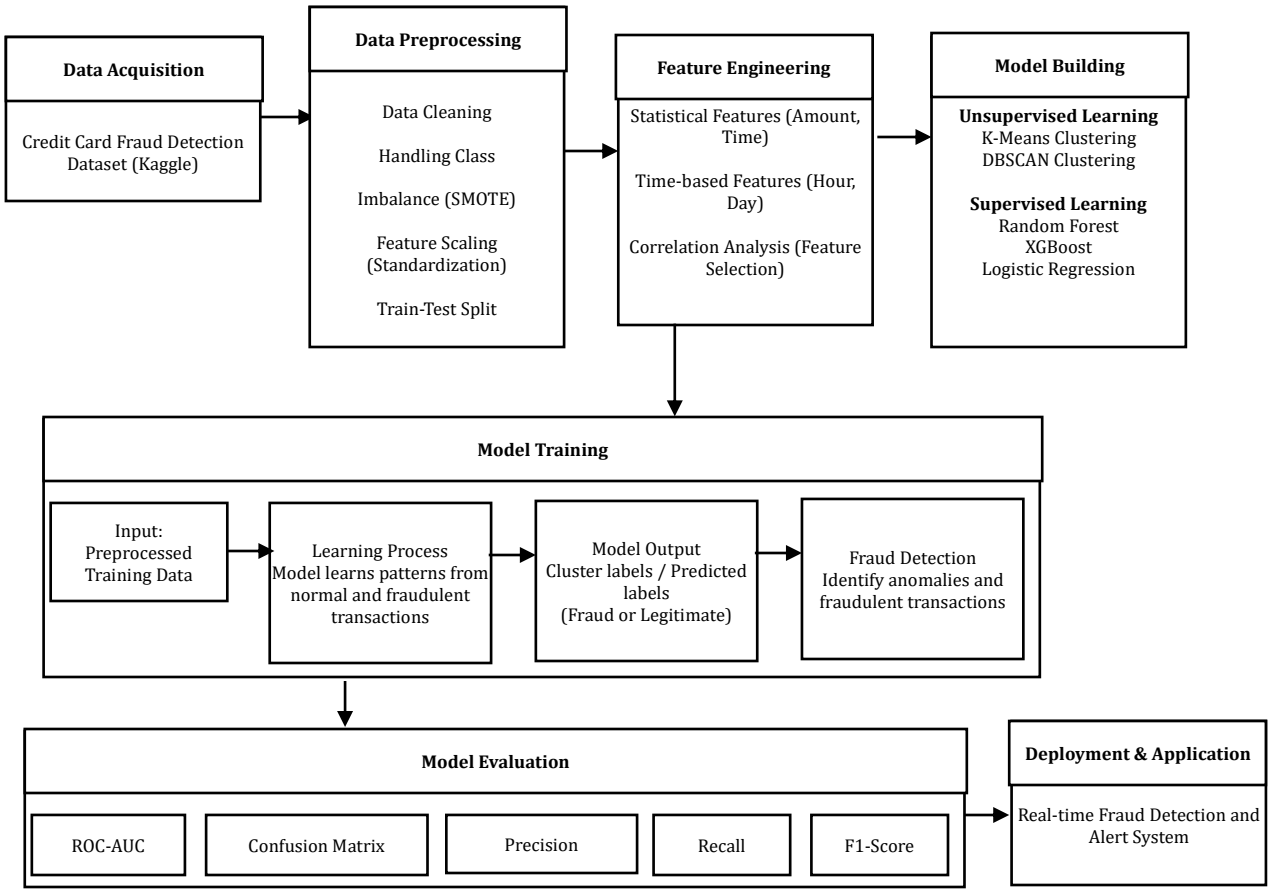

Figure 1 illustrates the flowchart of the proposed hybrid fraud detection framework. This process begins with data collection that involves anonymized credit card transactions. Preprocessing involves standardizing and normalizing the transaction data, together with addressing class imbalance issues. Feature extraction helps in obtaining important features of the transaction dataset. Building the models involves applying IF for the purpose of anomaly detection and applying DBSCAN to perform the clustering of suspicious transactions. In the training phase, patterns are obtained from the normal transactions and fraud.

Algorithm 1: Hybrid IF–DBSCAN Fraud Detection
Input:

- Transaction dataset $D = \{x_1, x_2, \dots, x_n\}$with features $[V1, V2, \dots, V28, Time, Amount]$
- IF parameters: number of estimators n_{trees}, contamination ratio τ
- DBSCAN parameters: epsilon ϵ, minimum points $minPts$

Output:

- Detected anomalous transactions A
- Clustered suspicious transaction groups $C = \{C_1, C_2, \dots, C_k\}$

Steps:

1. **Data Preprocessing:**

- Standardize and normalize features Time and Amount
- Verify no missing values exist
- Prepare dataset for unsupervised learning

2. **Anomaly Detection with** IF**:**

- Fit IF on preprocessed data $D$with $n_{trees}$and τ

- Compute anomaly score $s(x_i, n)$for each transaction x_i
- Select anomalous transactions:

$$A = \{x_i \in D \mid s(x_i, n) > \tau\}$$

3. Clustering with DBSCAN:

- Apply DBSCAN on $A$with parameters $\epsilon$and $minPts$
- Identify clusters $C = \{C_1, C_2, \ldots, C_k\}$
- Label transactions not in any cluster as noise (potential fraud)

4. Pattern Analysis:

- Inspect clusters $C$to uncover hidden patterns of fraudulent transactions
- Identify similarities in transaction amount, time, and PCA features within clusters

5. Performance Evaluation:

- Compute classification metrics
- Compute clustering metrics
- Compare hybrid model performance with IF alone

6. Output Results:

Return detected anomalies A, clusters C, and performance metrics

Algorithm 1 utilizes IF and DBSCAN to analyze patterns in financial transactions to detect fraud. In the first stage of fraud detection, the transactions are preprocessed through standardization and normalization of the Time and Amount features. Anomalies are then detected from this set of transactions through IF, which allocates an anomaly score for each transaction. Transactions that score higher on this anomaly score are flagged as fraudulent transactions. In the second stage, the anomalies are clustered through DBSCAN to discover dense groups of similar fraud behaviors. This step employs parameters such as epsilon (ϵ) and minimum points (minPts) to group the transactions into clusters and classify the anomalies as noise. Fraud patterns are discovered from these clusters, and the performance of this algorithm is evaluated based on classification measures and clustering.

3.5 Performance Metrics

To evaluate the hybrid fraud detection framework, both classification and clustering metrics are used:

Accuracy evaluates the proportion of properly identified transactions, as shown in equation (3):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Precision indicates the fraction of predicted fake transactions which are truly fake is shown in equation (4):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

Recall (Sensitivity) evaluates the fraction of actual fraudulent transactions correctly identified, as shown in equation (5):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

F1-score is the harmonic mean of precision and recall, which balances false positives and false negatives, and is shown in equation (6):

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Silhouette Score evaluates clustering quality by comparing intra-cluster cohesion and inter-cluster separation, as shown in equation (7):

$$S(i) = \frac{b(i) - a(i)}{max\big(a(i), b(i)\big)} \quad (7)$$

where $a(i)$ represents the mean distance of a point in the same cluster to other points in the same cluster, and $b(i)$represents the least mean distance to points in the closest cluster.

3.6 Implementation Details

Experiments were implemented in Python utilizing libraries such as scikit-learn, NumPy, and Pandas. Experiments were done on a computer that had an Intel Core i7 processor, 16 GB RAM, and an NVIDIA RTX graphics card. This

helped process transactional big data effectively. All hyperparameter optimization, model evaluation, and results reproducibility were done on this computing infrastructure.

4. Experimental Setup and Results

4.1 Split and Preprocessing of the Data

The Kaggle Credit Card Fraud dataset was employed, and divided into training and test datasets in a 80% to 20% ratio, ensuring adequate data for learning by the model, while reserving unseen data for evaluation purposes. Data preprocessing included normalization of the Time and Amount attributes, while the 28 PCA attributes (V1-V28) were normalized beforehand. There were no missing data entries in this dataset. The use of this preprocessing method ensured that the dataset was ready for use in an unsupervised environment and eliminated the impact of extremely high transaction amounts.

4.2 Comparative Evaluation

IF Model and the novel IF DBSCAN hybrid approach have been tested. In the IF model, anomalies are detected using the path length metric. Whereas the novel approach has used the IF for detecting anomalies and then has formed clusters from these anomalies. This helps in detecting any pattern related to suspicious activities. Performance evaluation is carried out on the basis of classification metrics.

Table 1: Performance Comparison of Fraud Detection Methods on Credit Card Transactions

References	Method(s)	Accuracy	Precision	Recall	F1-Score
[21]	IF, LOF	99.72%	NA	NA	NA
[22]	IF, LOF	NA	99.774% (IF) vs 99.65% (LOF)	27% (IF) vs 2% (LOF)	NA
[23]	IF, LOF, SVM, LR	99.6% (IF); 99.7% (LOF)	NA	NA	NA
[24]	IF, Logistic Regression	99.82% (train) / 74% (test)	0.49	0.49	0.49
[11]	IF, XGBoost, LOF, PCA	92%	92%	96%	NA
Proposed Study	IF + DBSCAN	99.88%	0.93	0.87	0.90

Table 1 provides a comparison of different fraud detection methods used in past literature, along with the newly suggested technique of hybrid IF and DBSCAN. Measures such as Recall, Accuracy, F1-Score, and Precision have been taken into consideration to analyze the performance of detection techniques. The hybrid IF and DBSCAN model shows better precision and F1-Score results, which means that this new model can successfully detect and isolate fraudulent transactions without compromising on accuracy.

Figure 2: Distribution of Anomaly Scores for Credit Card Transactions using IF

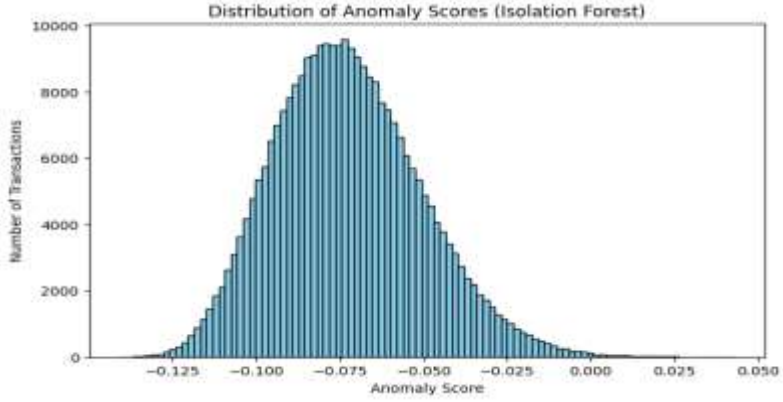

Figure 2 represents the distribution of anomaly scores based on the IF algorithm for the credit card transactions. The more the anomaly score, the greater the chances that the transaction is fraudulent. From the graph presented in Figure 2, it is clear how well the IF algorithm works to identify anomalies in an imbalanced dataset.

Table 2: Model Evaluation on Credit Card Fraud Dataset

Metric	**Hybrid** IF **+ DBSCAN**
Accuracy	*99.88%*
Precision	0.93
Recall	0.87
F1-Score	0.90
ROC-AUC	0.96
Silhouette Score	0.61
Davies–Bouldin Index	0.45

The Evaluation of the proposed model on the Kaggle Fraud dataset is provided in Table 2. As seen from the results obtained, the hybrid approach shows great efficiency in terms of fraud detection (accuracy = 99.88%), as well as provides a precision (0.93) and recall (0.87) values. Furthermore, the results show that the quality of clustering achieved is rather impressive, with the silhouette coefficient measuring up at 0.61 and a 0.45 Davies-Bouldin Index.

4.3 Statistical Validation and Performance Analysis

The superior performance of the Hybrid approach is proved through statistical analysis by employing p-value tests for F1 score, where $p < 0.05$ shows that there are significant differences in performance compared to IFs. In addition, ROC graphs prove better discrimination power with an increase in ROC-AUC from 0.92 to 0.96. Finally, high Silhouette scores (Silhouette score = 0.61) and low Davies-Bouldin Index (DBI = 0.45) values show that clusters formed using DBSCAN are well-defined and compact.

4.4 Ablation study

Table 3: Ablation Study of Each Component

Configuration	Accuracy	Precision	Recall	F1-Score
IF only	99.72%	0.85	0.77	0.81
DBSCAN on full dataset	99.65%	0.80	0.70	0.75
IF → DBSCAN (Hybrid)	**99.88%**	**0.93**	**0.87**	**0.90**

As depicted in Table 3, here is the contribution of each part in the hybrid method. As evident from the results obtained, it can be concluded that the integration of the IF method with the DBSCAN algorithm gives better performance in comparison to only using the DBSCAN algorithm or only the IF method.

5. Discussion

The findings of the experiments show that the presented model of the combination of the IF and DBSCAN algorithms allows identifying anomalous payments while forming meaningful clusters that help detect the hidden financial fraud schemes. As compared to the IF algorithm, the hybrid model shows better performance in terms of accuracy, precision, and F1-score values, emphasizing the effectiveness of reducing false alarms and more successful identification of rare fraud cases. Thanks to the formed clusters of anomalies, it is possible to identify specific financial transactions' similarities, like transaction amounts or merchant types, during a particular period.

One of the significant advantages of the model proposed is the fact that the approach allows scalability and adaptation to multidimensional, unbalanced data sets common in practical finance applications. Anomaly detection, along with density clustering, enables the detection of fraudulent cases regardless of whether such anomalies occur independently or form specific clusters. At the same time, some of the shortcomings of the proposed approach include the sensitivity to DBSCAN parameters ϵ and minPts, along with the impossibility of labeling borderline instances without additional research.

Insights generated through clustering can be effectively utilized by banks and other financial institutions. As suspicious activities tend to occur in clusters, managers can use these insights to target their efforts towards

investigating such clusters. In addition to that, since the framework enables making decisions about what should be monitored in real time, any anomalies in the data stream can instantly be picked up, and clustered patterns can be used to determine possible fraud attempts on the part of criminals.

6. Conclusion

The current study proposes an innovative solution that adopts the use of an amalgamation of IF and DBSCAN clustering to detect credit card fraud within extremely imbalanced transactional data sets. In the proposed algorithm, anomaly detection is initially achieved by the application of IF, followed by the clustering of the resulting anomalies through the application of DBSCAN. Through empirical analysis, the proposed method proved its superiority in detecting fraudulent activities within the Kaggle Fraud dataset. More specifically, it achieved at 99.88% accuracy, 0.93 precision, 0.87 recall, 0.90 F1-score, and 0.96 ROC-AUC. The effectiveness of the detected clusters of anomalies was evaluated using the Silhouette Score, which came up to 0.61, and the Davies–Bouldin Index, amounting to 0.45. While the proposed method enhances the accuracy of detecting rare instances of fraud, it also produces interpretable clusters, which would help banks in directing their efforts in investigating cases and managing risks. However, the research has certain drawbacks. First, the DBSCAN algorithm relies on the choice of parameters, which may necessitate regular training in case of dynamic patterns of fraud. Furthermore, the model has been tested on one dataset and might need adjustments in order to accommodate different banking systems globally. Some directions for further research include real-time analysis of transactions, implementation of a DL model, and evaluation of the proposed methodology on a multi-bank dataset. In conclusion, the research reveals that an ensemble of IF and DBSCAN algorithms can be an effective tool for detecting fraud using artificial intelligence technologies.

Declaration

Conflict of Interest
The authors declare no conflicts of interest regarding the publication of this paper.
Financial Statement
This research did not receive any specific funding or grants from public, commercial, or non-profit funding agencies.
Data Availability Statement
The dataset used in this study is publicly available from Kaggle and can be accessed at https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.

References

1. Herreros-Martínez, A., Magdalena-Benedicto, R., Vila-Francés, J., Serrano-López, A. J., Pérez-Díaz, S., & Martínez-Herráiz, J. J. (2025). Applied machine learning to anomaly detection in enterprise purchase processes: A hybrid approach using clustering and isolation forest. Information, 16(3), 177.
2. Maiti, A., Chakraborty, R., Basu, D., Sarkar, I., & Dutta, A. (2025). Unsupervised pattern discovery in cyber incidents using principal component analysis, K-means, DBSCAN, and isolation forest. In International Conference on Data Science and Network Engineering (pp. 314–324). Springer Nature Switzerland.
3. Lu, H. (2025). Evaluating the performance of SVM, isolation forest, and DBSCAN for anomaly detection. In ITM Web of Conferences (Vol. 70, Article 04012). EDP Sciences.
4. Fatlawi, H. K. (2025). Enhanced fraudulent detection using isolation forest and multi-cluster deep learning. Journal of Al-Qadisiyah for Computer Science and Mathematics, 17(1), 72–80.
5. Dhanesha, P., & Mehta, D. (2025). Uncovering hidden frauds: Isolation forest-based anomaly detection. In Intelligent Strategies for ICT: Proceedings of ICTCS 2024 (Vol. 4, p. 39).
6. Shyan, H., Singh, H., Yadav, N., Kumari, D., & Yashika, Y. (2025). Intelligent fraud detection in financial transactions using isolation forest and PCA. In 2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA) (pp. 1–7). IEEE.
7. Mandhala, V. N. (2025). Comparative analysis of clustering algorithms for financial fraud detection. International Journal of Safety & Security Engineering, 15(4).
8. Manoharan, G., Dharmaraj, A., Sheela, S. C., Naidu, K., Chavva, M., & Chaudhary, J. K. (2024). Machine learning-based real-time fraud detection in financial transactions. In 2024 Proceedings of the 3rd International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1–6). IEEE.

9. Anbazhagan, K. (2021). Anomaly detection in large-scale data using clustering and outlier analysis. International Journal of Advanced Research in Computer Science & Technology, 4(5), 5457–5461.
10. Udayakumar, R., et al. (2023). Integrated SVM-FFNN for fraud detection in banking financial transactions. Journal of Internet Services and Information Security, 13(4), 12–25.
11. Chapwanya, N., & Gorejena, K. N. (2025). Hybrid unsupervised machine learning for insurance fraud detection: PCA-XGBoost-LOF and isolation forest. Journal of Information Systems and Informatics, 7(1), 941–959.
12. Suganthi, V., & Jebathangam, J. (2025). A novel credit card fraud detection by outlier identification and elimination. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 16(3), 79–102. https://doi.org/10.58346/JOWUA.2025.I3.006
13. Ounacer, S., Ait El Bour, H., Oubrahim, Y., Ghoumar, M. Y., & Azzouazi, M. (2018). Using isolation forest in anomaly detection: The case of credit card transactions. Periodicals of Engineering and Natural Sciences, 6(2).
14. Agbeja, O., & Sokunle, O. T. (2019). An assessment of the role of external auditors in the detection and prevention of fraud in deposit money banks in Nigeria (2005–2014). International Academic Journal of Accounting and Financial Management, 6(1), 32–55. https://doi.org/10.9756/IAJAFM/V6I1/1910004
15. Almusallam, N., & Qayyum, J. (2025). A hybrid feature selection and clustering-based ensemble learning approach for real-time fraud detection in financial transactions. Computers, Materials & Continua, 85(2), 3653.
16. Kaliyaperumal, P., Periyasamy, S., Periyasamy, M., & Alagarsamy, A. (2024). Harnessing DBSCAN and autoencoder for hyper intrusion detection in cloud computing. Bulletin of Electrical Engineering and Informatics, 13(5), 3345–3354.
17. El Emary, I. M., Brzozowska, A., Popławski, Ł., Dziekański, P., & Glova, J. (2026). Anomaly detection in blockchain-based metaverse transactions using hybrid autoencoder and isolation forest models for risk identification and behavioral pattern analysis. International Journal of Research on Metaverse, 3(1), 46–63.
18. Alfaouri, F. T., Issa, H. B., Alnimer, R., & Althnaibat, O. H. A. (2026). Clustering techniques for discovering patterns in corporate law violations. Indian Journal of Information Sources and Services, 16(1), 169–178. https://doi.org/10.51983/IJISS-2026.16.1.18
19. Du, J., & Yu, B. (2023). Application of isolation forest algorithm in fraud detection of medical insurance big data. In 2023 8th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS) (Vol. 8, pp. 504–509). IEEE.
20. Sharma, A., & Menon, R. (2025). Blockchain integration in global accounting systems: Enhancing transparency and reducing fraud in cross-border transactions. Global Perspectives in Management, 3(3), 6–11.
21. Vijayakumar, V., Divya, N. S., Sarojini, P., & Sonika, K. (2020). Isolation forest and local outlier factor for credit card fraud detection system. International Journal of Engineering and Advanced Technology, 9, 261–265.
22. Rajeev, H., & Devi, U. (2022). Detection of credit card fraud using isolation forest algorithm. In Pervasive Computing and Social Networking: Proceedings of ICPCSN 2021 (pp. 23–34). Springer Nature Singapore.
23. Caroline Cynthia, P., & Thomas George, S. (2020). An outlier detection approach on credit card fraud detection using machine learning: A comparative analysis of supervised and unsupervised learning. In Intelligence in Big Data Technologies—Beyond the Hype: Proceedings of ICBDCC 2019 (pp. 125–135). Springer Singapore.
24. Akinola, K. E., Aina, D. A., Oyedele, O., & Braimoah, J. A. (2023). Credit card fraud detection using logistic regression and isolation forest algorithm. UNIZIK Journal of Engineering and Applied Sciences, 2(1), 187–195.