ANOMALY DETECTION: IMPROVING DIGITAL TRANSACTION FRAUD DETECTION WITH A HYBRID MODEL OF ISOLATION FOREST AND LONG SHORT-TERM MEMORY (LSTM)

Mutiara[1], Rendi Renaldi[2], Amelia Intan Sari[3], Angga Yunandar[4]

Informatics Engineering Study Program
Kaputama College of Informatics and Computer Management, Binjai City
Veteran Street No. 4A-9 A, Binjai City 25656 Binjai North Sumatra

E-mail: mutiaramutiara009@gmail.com, psrsatutandem@gmail.com, ameliaintan2001@gmail.com, anggayunandar@gmail.com

ABSTRACT

Financial transactions in the digital era are increasingly shifting to cashless methods due to their efficiency, but this has actually increased the risk of financial crime. This research aims to develop a more effective anomaly detection model for identifying suspicious transactions in digital transaction systems. This study presents a novel hybrid approach that combines the Isolation Forest Algorithm and Long Short-Term Memory (LSTM) to identify fraud. Isolation Forest was used to detect transactions that deviate from the normal user profile, while LSTM was used to detect various types of fraud. The results showed that out of 10,000 transaction data, the anomaly detection rate reached approximately 1.87% of the total transactions, with accuracy (100%), precision (99%), recall (100%), and F1-score (100%). The evaluation showed that this model was able to detect suspicious transactions with a very low error rate. The study concluded that this novelty is expected to contribute to the development of more accurate and efficient fraud detection models.

Keywords: Digital Era, Anomaly Detection, Digital Transactions, Isolation Forest, Long Short-Term Memory

Article History
Received: Januari 2026
Reviewed:Januari 2026
Published: Januari 2026

Plagiarism Checker No 235
Prefix DOI : 10.8734/Kohesi.v1i2.365
Copyright : Author
Publish by : Kohesi

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

1. Introduction

The rapid development of digital technology has driven an increase in online financial transactions. The ease and speed of these digital transactions provide convenience for users. However, on the other hand, this increase in digital transaction activity is also accompanied by an increased risk of fraud. Fraudsters continue to develop new tactics to exploit system vulnerabilities and harm users. Therefore, developing effective fraud detection methods is crucial.

Anomaly Detection is the process of detecting data that deviates significantly from common data patterns. The application of AD to digital transaction data has yielded numerous applications, such as identifying activity that deviates from normal patterns. Activity considered anomalous often indicates potential fraud. Therefore, developing effective anomaly detection methods is crucial to protecting transaction security and maintaining consumer trust.

Traditional fraud detection methods are often limited in addressing complex and rapidly evolving fraud patterns. Rule-based algorithms, for example, struggle to adapt to changing fraudster behavior. Therefore,

more sophisticated approaches are needed to address these challenges. Artificial intelligence (AI), particularly machine learning techniques such as Isolation Forest and Long Short-Term Memory (LSTM), offer significant potential in detecting anomalies in financial transactions.

Thus, this study proposes a hybrid model that combines both models to improve detection accuracy. By combining these two algorithms, it is hoped that a model can be produced that is more capable of detecting various types of fraud, both outliers and those involving gradual changes in transaction patterns.

2. Materials and Methods

2.1. Research Approach

This research uses a quantitative approach with the concept of digital transaction data analysis that involves collecting numerical data, then processing and analyzing it using the Isolation Forest method, Long Short-Term Memory, then combining the two algorithms to determine which is more accurate in finding transactions that deviate from normal patterns, and providing early warnings if an anomaly is detected.

2.2. Research Questions

The research problem formulation is based on the following research objectives:

• RQ1: What features are most relevant for fraud detection in digital transactions?

• RQ2: How can the Isolation Forest and Long Short-Term Memory algorithms be effectively combined?

2.3. Research Design and Objects

This research is designed to detect fraud in digital transaction data on E-commerce platform X, the research object.

- Data collection: Using synthetic data with Python and Numpy features
- Data preprocessing: Cleaning and transformation using the Python programming language. Anaconda Navigator application with feature selection, Pandas, Numpy, and Matplotlib.
- Model building: Creating Isolation Forest, LSTM, and hybrid models using Scikit-learn and TensorFlow.
- Model training: Training the model using optimization algorithms using TensorFlow.
- Results analysis: Calculating evaluation metrics (accuracy, precision, recall, and F1-score) using Scikit-learn.
- Comparative analysis: Comparing the performance of the hybrid model with the baseline Python and Matplotlib models.

2.4. Data Collection

Data source: The data source to be used is a dataset of 10,000 transaction data, which includes key features such as transaction date (transaction_date), customer ID (customer_id), transaction amount (amount), and transaction status (status). Synthetic data containing: Transaction data, customer ID, payment information, nominal, and payment status.

2.5. Data Pre-processing

- Data cleansing: Eliminating inconsistent, duplicate, or erroneous data.
- Data transformation: Converting data into a format suitable for analysis, for example, converting categorical data to numeric (e.g., using one-hot encoding).

- Data normalization: Scaling data to have the same range of values, making it easier to compare.

2.6. Model Development

- Isolation Forest Algorithm: Identifies data that "doesn't match" or deviates from common patterns in the data. In the transaction context, this "mismatch" data is often an indication of fraud.
- LSTM: Processes transaction data to learn complex patterns in historical transaction data, then predicts whether a new transaction is fraudulent or not.
- Hybrid Model: The hybrid Isolation Forest and LSTM method offers a comprehensive approach to digital transaction fraud detection. By combining the strengths of both algorithms, we can build a more accurate and robust model for detecting various types of fraud.

2.7. Model Training

Isolation Forest Model Training:

- Isolation Forest is trained on transaction data to identify transactions that significantly deviate from normal patterns.
- This model generates an anomaly score for each transaction.

LSTM Model Training:

- Transaction data, enriched with anomaly scores from Isolation Forest, is used to train an LSTM model.
- The LSTM learns temporal patterns in the data, including interactions between normal transactions and those deemed anomalous by Isolation Forest.

Prediction:

- The trained LSTM model is used to predict whether a new transaction is fraudulent.

2.8. Assessment Criteria

Assessment criteria for processed data results:

- Precision: This measures the proportion of correct positive predictions. High precision means very few normal transactions are misclassified as anomalies.
- Recall: This measures the proportion of actual anomalies successfully identified by the model. High recall means very few anomalies are missed.
- F1-score: This is the harmonic mean between precision and recall. A high F1-score indicates that the model has both high precision and recall.

2.9. System Design

The design that will be developed in this research can be seen in
Figure 1. Flowchart

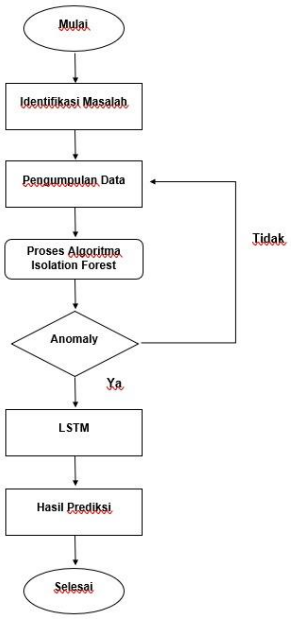


Figure 1. Flowchart

3. Result and Discussion

3.1. Result

The results of data processing using the Isolation Forest Algorithm can be seen in Figure 2 and Figure 3.

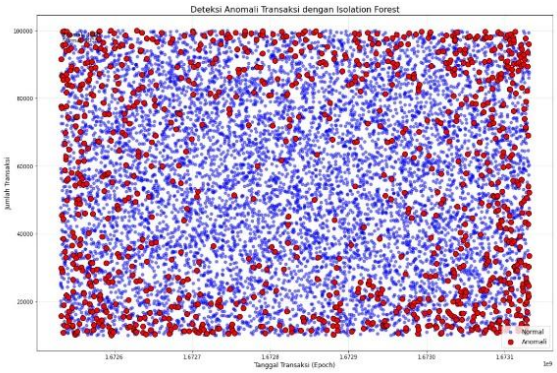


Figure 2. Transaction anomaly graph with Isolation Forest

```
Laporan Deteksi Anomali Isolation Forest:
Total Transaksi: 10000
Jumlah Anomali: 1000
Persentase Anomali: 10.00%

Contoh Transaksi Anomali:
    transaction_id    transaction_date  customer_id        amount     status
9               10 2023-01-01 00:09:00         1121  15749.179069  cancelled
11              12 2023-01-01 00:11:00         1214  29828.718876    pending
14              15 2023-01-01 00:14:00         1087  77659.370891    pending
15              16 2023-01-01 00:15:00         1372  91561.016749  cancelled
21              22 2023-01-01 00:21:00         1308  45468.113750  cancelled

Distribusi Anomali per Status:
cancelled    62.9
pending      35.4
completed     1.7
Name: status, dtype: float64

Statistik Deskriptif Transaksi Anomali:
              amount
count    1000.000000
mean    53624.226740
std     31890.659426
min     10114.126922
25%     21368.787863
50%     51234.828922
75%     86261.100843
max     99986.432004
```

Figure 3. Table of results of Isolation Forest Anomaly Detection

Based on the anomaly detection generated from Isolation Forest, the data is then processed using the Long Short-Term Memory model, which can be seen in Figure 4 and Figure 5.

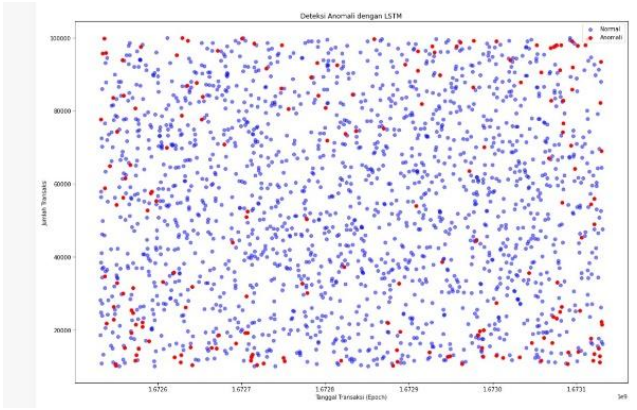
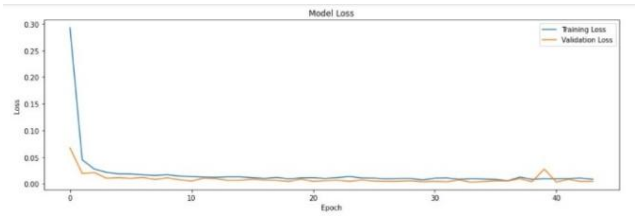


Figure 4. Transaction anomaly graph with LSTM




```
Jumlah Anomali Terdeteksi LSTM (Test Set): 188
```

Figure 5. Graph of LSTM Anomaly Detection results

3.2. Discussion Results

The results show that a hybrid model combining Isolation Forest and LSTM performs better in detecting digital transaction fraud than either model alone. This demonstrates that the combination of the two algorithms can capture both static and dynamic anomalies. Isolation Forest is effective in detecting transactions that deviate from normal patterns, while LSTM is capable of capturing complex temporal patterns.

The evaluation of the results of combining the two algorithms can be seen in Figure 6.

```
Confusion Matrix:
[[1812    1]
 [   0  187]]

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      1813
           1       0.99      1.00      1.00       187

    accuracy                           1.00      2000
   macro avg       1.00      1.00      1.00      2000
weighted avg       1.00      1.00      1.00      2000
```

Figure 6. Evaluation Table of the Results of Combining Isolation Forest and Long Short-Term Memory (LSTM)

The model evaluation results using commonly used metrics in anomaly detection performance testing, namely accuracy (100%), precision (99%), recall (100%), and F1-score (100%), showed very satisfactory performance. With an anomaly detection rate reaching approximately 1.87% of total transactions, this model is able to detect suspicious transactions with a very low error rate, ensuring high validity of the detection results. These metrics also indicate that this model is very effective in detecting fraud on large datasets, where only a small number of transactions are anomalous.

Figure 1 shows a graph of the change in loss during training, which shows a significant reduction in loss on both training and validation data.



Figure 1: Training Loss vs Validation Loss Graph

Figure 2 shows the model accuracy continuously increasing, which indicates that the model is successfully learning from the data.



Figure 2: Training Accuracy vs Validation Accuracy graph

ROC AUC (Receiver Operating Characteristic - Area Under the Curve)

- ROC Curve: This is a graph that illustrates the performance of a classification model by varying the classification threshold. The x-axis represents the false positive rate (FPR), which is the proportion of negative examples incorrectly classified as positive. The y-axis represents the true positive rate (TPR) or recall, which is the proportion of positive examples correctly classified as positive.
- AUC: The area under the ROC curve. The larger the area under the curve, the better the model's performance. The AUC provides an overall picture of the model's ability to distinguish between positive and negative classes.

3.3. Research Novelty

The novelty of this research lies in the combination of two algorithms: Isolation Forest and Long Short-Term Memory (LSTM) in handling digital transaction fraud cases. The use of a hybrid method, combining the

isolation forest algorithm with long short-term memory (LSTM), to detect anomalies in transaction data improves data validity accuracy and process efficiency. This combination produces a significant new approach to fraud detection in digital transactions.

This study successfully developed and evaluated a hybrid model that combines two leading anomaly detection algorithms, namely Isolation Forest and Long Short-Term Memory (LSTM), to detect fraud in digital transactions. This model was tested using a transaction dataset consisting of 10,000 transaction data, which includes key features such as transaction date (transaction_date), customer ID (customer_id), transaction amount (amount), and transaction status (status). Based on the experiments conducted, the results obtained showed that the combination of these two algorithms provides significantly superior detection performance compared to when the algorithms are applied separately.

The main contribution of this research is the development of a more accurate and efficient fraud detection model, which uses a hybrid approach based on the Isolation Forest and LSTM algorithms to detect digital transaction anomalies. The application of this model in the industrial world can provide a more effective solution to address the growing risk of fraud in the digital financial world. Thus, this research not only enriches the scientific literature in the field of anomaly detection but also offers relevant contributions to the development of digital transaction security systems.

4. Conclusion

The conclusions in this study are:

1. Differences between the Isolation Forest and Long Short-Term Memory Algorithms
   - Isolation Forest demonstrates its ability to effectively detect anomalies, especially in transactions that deviate from normal patterns, while
   - LSTM provides advantages in identifying more complex temporal patterns, where transaction data contains important time relationships.

2. Advantages of Using a Hybrid Model
   - The advantage of this hybrid model lies in its ability to capture both static and dynamic anomalies, which cannot be detected using a single algorithm.

3. Evaluation Results
   - Model evaluation results using the metrics used in anomaly detection performance testing, namely accuracy (100%), precision (99%), recall (100%), and F1-score (100%), demonstrated very satisfactory performance.
   - With an anomaly detection rate of approximately 1.87% of total transactions, this model is capable of detecting suspicious transactions with a very low error rate, ensuring high validity of the detection results.

4. Research Novelty
   - This research contributes to the development of a more accurate and efficient fraud detection model, which uses a hybrid approach based on the Isolation Forest and LSTM algorithms to detect digital transaction anomalies.

5. General Conclusions

- Overall, the results of this study confirm that the use of a hybrid method combining Isolation Forest and LSTM can make a significant contribution to the development of more sophisticated, accurate, and efficient fraud detection systems.

5. References

[1] Ayan Chatterjee, Bestoun S. Ahmed, "IoT anomaly detection methods and applications: A survey" *Department of Mathematics and Computer Science, Karlstad University, 651 88 Karlstad, Sweden*, Internet of Things 19 (2022) 100568. [Online]. Available: https://doi.org/10.1016/j.iot.2022.100568

[2] Ngoc-Thanh Nguyen, Rogardt Heldal, Patrizio Pelliccione, "Concept-drift-adaptive anomaly detector for marine sensor data streams" *Western Norway University of Applied Sciences, Norway,* Internet of Things 28 (2024) 101414. [Online]. Available: https://doi.org/10.1016/j.iot.2024.101414

[3] Marian Stan, Adriana Dima, Dag Øivind Madsen, Cosmin Dobrin, "Quantifying impact: Bibliometric examination of IoT's evolution in sustainable development" *Bucharest University of Economic Studies, 010374, Bucharest, Romania,* Internet of Things 28 (2024) 101370. [Online]. Available: https://doi.org/10.1016/j.iot.2024.101370

[4] Zainab Alwaisi, Tanesh Kumar, Erkki Harjula, Simone Soderi, "Securing constrained IoT systems: A lightweight machine learning approach for anomaly detection and prevention" *School of Electrical Engineering, Aalto University, Finland,* Internet of Things 28 (2024) 101398. [Online]. Available: https://doi.org/10.1016/j.iot.2024.101398

[5] Mahdi Mnif, Salwa Sahnoun, Yasmine Ben Saad, Ahmed Fakhfakh, Olfa Kanoun, "Combinative model compression approach for enhancing 1D CNN efficiency for EIT-based Hand Gesture Recognition on IoT edge devices" *a National School of Electronics and Telecommunications of Sfax, University of Sfax, Tunisia,* Internet of Things 28 (2024) 101403. [Online]. Available: https://doi.org/10.1016/j.iot.2024.101403

[6] Nicolas Farabegoli, Danilo Pianini, Roberto Casadei, Mirko Viroli, "Dynamic IoT deployment reconfiguration: A global-level self-organisation approach" *Alma Mater Studiorum – University of Bologna, Cesena, Italy,* Internet of Things 28 (2024) 101412. [Online]. Available: https://doi.org/10.1016/j.iot.2024.101412