KAMPUS AKADEMIK PUBLISING
Jurnal Sains Student Research (JSSR)
Vol.4, No.5 Oktober 2026
e-ISSN: 3025-9851, p-ISSN : 3025-986X, Hal 1337-1354
DOI : https://doi.org/10.61722/jssr.v4i5.12743

DETEKSI PENIPUAN TRANSAKSI PADA PLATFORM MARKETPLACE MENGGUNAKAN ALGORITMA ISOLATION FOREST

Rakha Aryagunawibawa[1*], Puspita Nurul Sabrina[2], Herdi Ashaury[3]
[1,2,3] Studi Teknik Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani
*Penulis Korespondensi: rakhaarya57@gmail.com, puspita.sabrina@lecture.unjani.ac.id, herdiashaury@lecture.unjani.ac.id

Abstract. *The increasing use of marketplace platforms has raised the risk of transaction fraud. Fraud detection is challenging because fraudulent transactions are far less frequent than legitimate transactions, creating an imbalanced dataset. This study analyzes the capability of Isolation Forest for fraud detection, evaluates the influence of preprocessing and hyperparameter settings, and compares its performance with Local Outlier Factor (LOF). The experiment uses the public Fraudulent E-Commerce Transactions dataset from Kaggle containing 1,048,575 transactions: 996,055 legitimate transactions (94.99%) and 52,520 fraudulent transactions (5.01%). The workflow includes data cleaning, feature selection, feature engineering, One-Hot Encoding, StandardScaler, an 80:20 stratified train-test split, baseline modeling, hyperparameter tuning, threshold tuning, evaluation, and comparison with LOF. Isolation Forest is trained in an unsupervised manner, while the Is Fraudulent label is used only as evaluation ground truth. The final configuration uses contamination 0.05, 250 estimators, max_samples 0.80, threshold 0.45, and random_state 42. The model achieves Accuracy 0.9820, Precision 0.8201, Recall 0.8200, F1-Score 0.8200, ROC-AUC 0.9187, and PR-AUC 0.6458. The confusion matrix contains 197,321 True Negatives, 8,613 True Positives, 1,890 False Positives, and 1,891 False Negatives. Isolation Forest outperforms LOF on most metrics, although LOF provides slightly higher Recall. The detection result is intended as an early indicator for transactions requiring further review.*

Keywords: *Fraud Detection, Marketplace, Isolation Forest, Anomaly Detection, Local Outlier Factor, Imbalanced Dataset.*

Abstrak. Peningkatan penggunaan platform marketplace turut meningkatkan risiko penipuan transaksi (fraud). Deteksi fraud menjadi tantangan karena jumlah transaksi fraud jauh lebih sedikit dibandingkan transaksi normal sehingga membentuk distribusi kelas yang tidak seimbang. Penelitian ini bertujuan menganalisis kemampuan algoritma Isolation Forest dalam mendeteksi transaksi fraud, mengevaluasi pengaruh preprocessing dan konfigurasi hyperparameter, serta membandingkan performanya dengan Local Outlier Factor (LOF). Penelitian menggunakan pendekatan kuantitatif melalui eksperimen data mining pada dataset Fraudulent E-Commerce Transactions dari Kaggle yang berjumlah 1.048.575 transaksi, terdiri atas 996.055 transaksi normal (94,99%) dan 52.520 transaksi fraud (5,01%). Tahapan penelitian meliputi data cleaning, seleksi atribut, feature engineering, One-Hot Encoding, StandardScaler, stratified train-test split 80:20, baseline model, hyperparameter tuning, threshold tuning, evaluasi, dan perbandingan dengan LOF. Isolation Forest diterapkan secara unsupervised; label Is Fraudulent hanya digunakan sebagai ground truth evaluasi. Skenario preprocessing lengkap menghasilkan keseimbangan performa lebih baik. Konfigurasi akhir menggunakan contamination 0,05, n_estimators 250, max_samples 0,80, threshold 0,45, dan random_state 42. Model menghasilkan Accuracy 0,9820, Precision 0,8201, Recall 0,8200, F1-Score 0,8200, ROC-AUC 0,9187, dan PR-AUC 0,6458. Confusion Matrix menunjukkan 197.321 True Negative, 8.613 True Positive, 1.890 False Positive, dan 1.891 False Negative. Dibandingkan LOF, Isolation Forest unggul pada sebagian besar metrik, sedangkan LOF memiliki Recall sedikit lebih tinggi. Hasil deteksi diposisikan sebagai indikasi awal transaksi mencurigakan untuk pemeriksaan lebih lanjut.

Kata Kunci: fraud detection, marketplace, Isolation Forest, anomaly detection, Local Outlier Factor, imbalanced dataset.

Naskah Masuk:01 September 2026; Revisi: 02 September 2026; Diterima: 03 September 2026; Terbit: 19 September 2026

1. LATAR BELAKANG

Perkembangan marketplace meningkatkan jumlah transaksi digital dan mempermudah aktivitas jual beli daring, tetapi pertumbuhan volume transaksi juga memperbesar risiko transaksi fraud yang semakin sulit dideteksi secara manual [1]–[3]. Fraud dapat muncul dalam bentuk akun palsu, transaksi fiktif, penyalahgunaan metode pembayaran, maupun manipulasi pola transaksi [1], [4]. Tantangan utama muncul karena transaksi fraud sering menyerupai transaksi normal dan hanya memperlihatkan penyimpangan pada kondisi tertentu [3], [5].

Dataset transaksi fraud pada umumnya berukuran besar dan memiliki distribusi kelas tidak seimbang. Dataset yang digunakan pada penelitian ini berisi 1.048.575 transaksi, dengan 52.520 transaksi fraud (5,01%) dan 996.055 transaksi normal (94,99%). Dominasi kelas normal membuat proses identifikasi kelas minoritas menjadi lebih sulit dan menyebabkan Accuracy saja tidak memadai untuk menilai kinerja sistem [6], [7].

Pendekatan anomaly detection berbasis machine learning dipilih untuk mengenali transaksi yang menyimpang dari pola mayoritas. Isolation Forest mengisolasi data melalui partisi acak; data anomali cenderung lebih cepat terisolasi dibandingkan data normal [8], [9]. Karakteristik ini relevan untuk fraud detection karena model dapat bekerja secara unsupervised dan tidak membutuhkan label sebagai target pembelajaran. Label fraud pada penelitian ini digunakan hanya pada tahap evaluasi sebagai ground truth.

Penelitian terdahulu telah membahas fraud detection menggunakan supervised learning, hybrid anomaly detection, Isolation Forest, serta LOF. Namun, pembahasan mengenai pengaruh preprocessing, sensitivitas contamination, n_estimators, max_samples, dan threshold pada dataset marketplace dengan fitur numerik-kategorikal masih terbatas [10], [12]–[14]. Karena itu, penelitian ini tidak hanya menilai hasil akhir Isolation Forest, tetapi juga mengevaluasi pengaruh tahapan pengolahan data dan konfigurasi parameter terhadap performa model.

Kontribusi penelitian diarahkan pada tiga aspek. Pertama, penelitian menggunakan dataset marketplace dengan karakteristik fitur campuran, berbeda dari banyak penelitian credit card fraud yang didominasi fitur numerik hasil transformasi. Kedua, dilakukan eksperimen dua skenario preprocessing serta hyperparameter dan threshold tuning. Ketiga, evaluasi mencakup Precision, Recall, F1-Score, Accuracy, ROC-AUC, PR-AUC, False Positive, dan False Negative serta perbandingan dengan LOF. Tujuannya adalah

memperoleh gambaran yang lebih menyeluruh mengenai efektivitas Isolation Forest pada data transaksi marketplace yang imbalanced.

2. KAJIAN TEORITIS

Penelitian Terdahulu dan Kesenjangan Penelitian

Kajian terdahulu menunjukkan bahwa fraud detection pada transaksi digital telah ditangani menggunakan supervised learning, metode hybrid, dan anomaly detection. Ringkasan berikut dipilih karena paling berkaitan dengan posisi penelitian, yaitu pemanfaatan Isolation Forest, penanganan imbalanced data, konteks e-commerce, dan perbandingan dengan LOF.

Tabel 1. Ringkasan Penelitian Terdahulu yang Relevan

Penelitian	Metode / Fokus	Temuan Utama	Perbedaan dengan Penelitian Ini
Ismanda et al. (2025) [12]	Hybrid Isolation Forest + K-Means	Menggabungkan klasterisasi awal dan Isolation Forest untuk meningkatkan deteksi pada data imbalanced.	Penelitian ini menggunakan Isolation Forest secara mandiri dan menelaah preprocessing, hyperparameter, threshold, FP dan FN.
Waspada et al. [11]	Isolation Forest pada credit card fraud	Menganalisis performa Isolation Forest dengan ROC-AUC/AUCPR dan menunjukkan relevansinya pada data highly imbalanced.	Penelitian ini menggunakan dataset marketplace dengan fitur campuran serta evaluasi preprocessing dan parameter.
Zhang et al. (2022) [10]	Optimized anomaly detection / Isolation Forest	Mengusulkan optimasi untuk menangani class imbalance pada fraud detection.	Penelitian ini tidak menggunakan sampling dan menekankan feature engineering serta konfigurasi model.
Porwal & Mukund (2019) [13]	Outlier detection pada e-commerce	Mengkaji deteksi outlier pada data e-commerce berskala besar.	Penelitian ini memfokuskan Isolation Forest pada dataset marketplace dan menguji pengaruh preprocessing serta parameter.

Penelitian	Metode / Fokus	Temuan Utama	Perbedaan dengan Penelitian Ini
John & Naaz (2019) [14]	Perbandingan Isolation Forest dan LOF	Membandingkan akurasi dan waktu komputasi; Isolation Forest dilaporkan lebih cepat dan akurat daripada LOF.	Penelitian ini menambahkan tuning, evaluasi ROC-AUC/PR-AUC, dan analisis kesalahan prediksi.

Berdasarkan studi tersebut, penelitian ini mengidentifikasi tiga kesenjangan. Kesenjangan metodologi muncul karena sebagian penelitian menggunakan pendekatan hybrid atau deep learning yang lebih kompleks, sedangkan pengaruh preprocessing dan konfigurasi parameter Isolation Forest belum banyak dibahas pada konteks marketplace [12], [18]. Kesenjangan objek data berkaitan dengan dominasi dataset kartu kredit yang banyak memuat fitur numerik/PCA; penelitian ini menggunakan data marketplace dengan atribut numerik dan kategorikal [11], [14]. Kesenjangan evaluasi muncul karena sebagian studi menekankan Accuracy, sementara penelitian ini menganalisis trade-off False Positive dan False Negative serta metrik yang lebih sesuai untuk imbalanced data [4], [13].

Isolation Forest dan Anomaly Score

Isolation Forest membangun sekumpulan Isolation Tree (iTree). Pada setiap pohon, satu fitur dan nilai split dipilih secara acak. Data anomali yang memiliki karakteristik berbeda dari mayoritas cenderung membutuhkan lebih sedikit pemisahan sehingga memiliki path length lebih pendek [8], [19]. Perbedaan path length digunakan untuk membentuk anomaly score.

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

Pada persamaan tersebut, s(x,n) adalah anomaly score, E(h(x)) adalah rata-rata panjang jalur data x pada seluruh Isolation Tree, sedangkan c(n) merupakan faktor normalisasi berdasarkan jumlah sampel. Nilai anomaly score yang semakin mendekati 1 menunjukkan data semakin anomali, sedangkan skor yang semakin rendah menunjukkan karakteristik yang lebih mendekati transaksi normal [8], [19], [26]. Pada implementasi penelitian, anomaly score dinormalisasi ke rentang 0–1 menggunakan MinMaxScaler untuk memudahkan interpretasi dan threshold tuning.

Local Outlier Factor dan Metrik Evaluasi

Local Outlier Factor (LOF) merupakan anomaly detection berbasis kepadatan yang menilai keanehan suatu observasi berdasarkan kepadatan lokal terhadap tetangga terdekat. Berbeda dari Isolation Forest yang melakukan isolasi melalui partisi acak, LOF bergantung pada perhitungan jarak dan lebih sensitif terhadap skala fitur sehingga standardisasi penting pada tahap perbandingan [14].

Evaluasi menggunakan Confusion Matrix yang menghasilkan True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Precision mengukur ketepatan prediksi fraud, Recall mengukur kemampuan menemukan seluruh transaksi fraud, dan F1-Score mengukur keseimbangan Precision dan Recall. Accuracy tetap dilaporkan tetapi tidak digunakan sebagai satu-satunya indikator karena dataset bersifat imbalanced [6], [20], [21].

Tabel 2. Rumus Metrik Evaluasi Utama

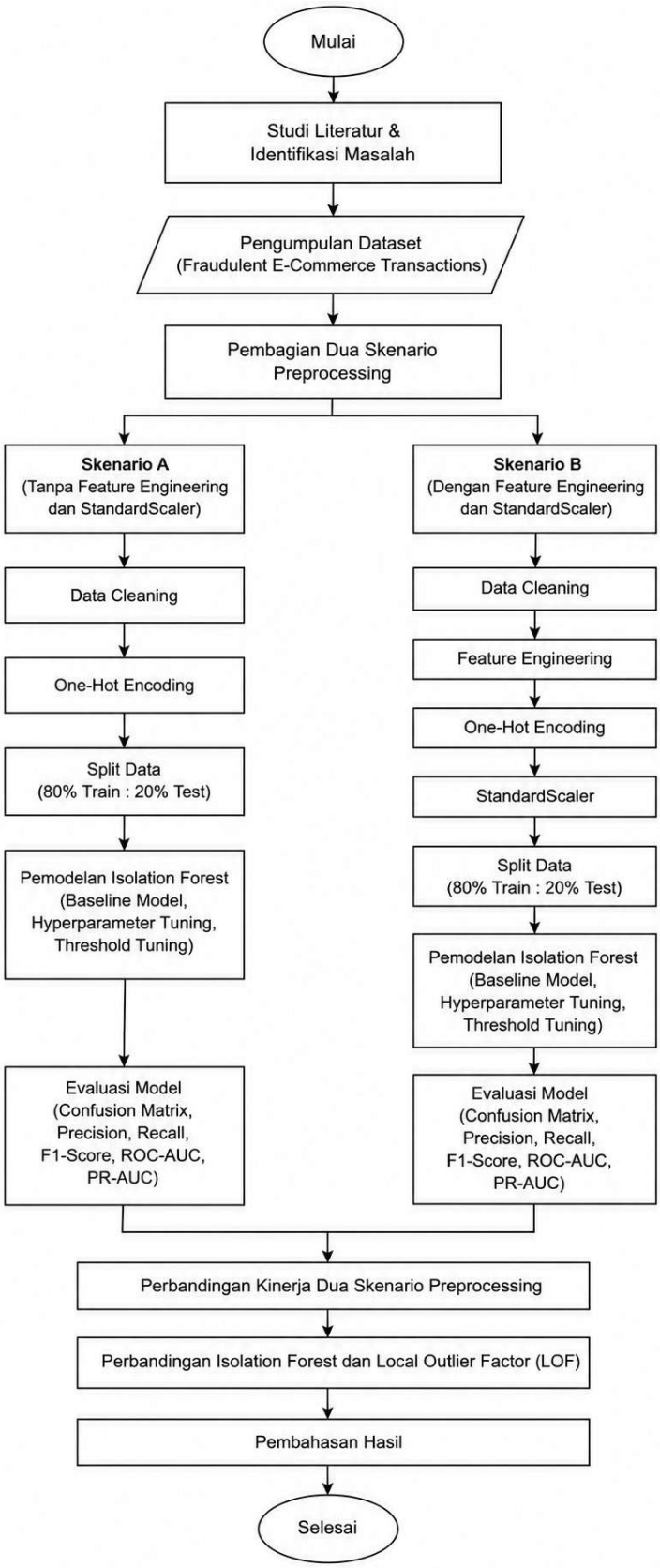
Metrik	Rumus
Precision	$Precision = \frac{TP}{TP + FP}$
Recall	$Recall = \frac{TP}{TP + FN}$
F1-Score	$F1\ Score = 2\ \times \frac{Precision \times Recall}{Precision + Recall}$
Accuracy	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$

ROC-AUC digunakan untuk menilai kemampuan diskriminasi model pada berbagai threshold, sedangkan PR-AUC menekankan hubungan Precision–Recall dan lebih informatif pada dataset tidak seimbang [22], [23]. Oleh karena itu, kedua metrik digunakan bersama metrik berbasis Confusion Matrix.

3. METODE PENELITIAN

Pendekatan dan Desain Eksperimen

Penelitian menerapkan pendekatan kuantitatif dengan eksperimen data mining. Variabel yang dimanipulasi meliputi tahapan preprocessing dan konfigurasi parameter, sedangkan variabel terikat adalah kinerja model dalam mendeteksi transaksi fraud. Alur penelitian dimulai dari studi literatur, pengumpulan dataset, dua skenario preprocessing, pembagian data, pemodelan Isolation Forest, tuning, evaluasi, dan perbandingan dengan LOF.



Gambar 1. Flowchart Penelitian

Dataset dan Pembagian Data

Dataset sekunder Fraudulent E-Commerce Transactions diperoleh dari repositori publik Kaggle. Dataset memuat informasi transaksi, profil pelanggan, metode pembayaran, kategori produk, perangkat, waktu transaksi, usia akun, dan label Is Fraudulent. Seluruh data yang lolos pemeriksaan digunakan sebagai unit analisis; tidak dilakukan sampling dari dataset. Label fraud tidak digunakan sebagai fitur input pembelajaran Isolation Forest, tetapi hanya sebagai acuan evaluasi.

Tabel 3. Distribusi Kelas Dataset

Kelas	Jumlah	Persentase
Normal (0)	996.055	94,99%
Fraud (1)	52.520	5,01%
Total	1.048.575	100%

Tabel 4. Atribut Utama Dataset Sebelum Seleksi

Atribut	Tipe Data	Keterangan
Transaction ID	String	Kode unik setiap transaksi.
Customer ID	String	Kode unik pelanggan.
Transaction Amount	Float	Nominal transaksi.
Transaction Date	Datetime	Tanggal dan waktu transaksi.
Payment Method	Kategorikal	Metode pembayaran.
Product Category	Kategorikal	Kategori produk.
Quantity	Int	Jumlah barang yang dibeli.
Customer Age	Int	Usia pelanggan.
Customer Location	String	Lokasi pelanggan.
Device Used	Kategorikal	Perangkat yang digunakan.
IP Address	String	Alamat protokol internet pengguna.
Account Age Days	Int	Usia akun pelanggan dalam hari.
Transaction Hour	Int	Jam transaksi (0–23).
Shipping Address	String	Alamat tujuan pengiriman.
Billing Address	String	Alamat penagihan.
Is Fraudulent	Boolean (0/1)	Label evaluasi: normal atau fraud.

Dari 16 atribut awal, atribut identitas atau ber-kardinalitas tinggi tidak digunakan langsung dalam pemodelan. Setelah feature engineering dan One-Hot Encoding, Skenario B menghasilkan 22 kolom, dengan 21 fitur input dan satu label Is Fraudulent untuk evaluasi.

Data dibagi menggunakan stratified train-test split 80:20 agar proporsi kelas normal dan fraud tetap sama pada data training dan testing. Hasil pembagian terdiri atas 838.860 data training dan 209.715 data testing. Proporsi kelas pada kedua subset tetap sekitar 94,99% normal dan 5,01% fraud.

Preprocessing dan Skenario Eksperimen

Preprocessing meliputi data cleaning, seleksi atribut, feature engineering, One-Hot Encoding, dan StandardScaler. Pemeriksaan awal terhadap 1.048.575 transaksi tidak menemukan missing value maupun data duplikat. Atribut dengan kardinalitas tinggi atau fungsi identitas, seperti Transaction ID, Customer ID, IP Address, Shipping Address, Billing Address, dan Customer Location, tidak digunakan secara langsung. Transaction Date diolah melalui feature engineering menjadi Day, Month, Weekday, dan Weekend. Atribut kategorikal Payment Method, Product Category, dan Device Used ditransformasikan menggunakan One-Hot Encoding.

Tabel 5. Detail Dua Skenario Preprocessing

Komponen	Skenario A	Skenario B
Dataset	Cleaning + OHE	Cleaning + FE + OHE + StandardScaler
Feature Engineering	Tidak	Ya
One-Hot Encoding	Ya	Ya
StandardScaler	Tidak	Ya
Algoritma	Isolation Forest	Isolation Forest
Training / Testing	80% / 20%	80% / 20%
Kolom dataset	18	22
Fitur input model	17	21
Metrik	Precision, Recall, F1, Accuracy	Precision, Recall, F1, Accuracy

Skenario A digunakan sebagai preprocessing yang lebih sederhana, yaitu data cleaning dan One-Hot Encoding. Skenario B menggunakan seluruh tahapan yang dirancang: data cleaning, feature engineering, One-Hot Encoding, dan StandardScaler. Karena feature engineering dan StandardScaler diterapkan bersama pada Skenario B, penelitian tidak mengisolasi kontribusi masing-masing tahap secara individual.

Pemodelan, Tuning, dan Perbandingan

Pemodelan diawali dengan baseline Isolation Forest menggunakan konfigurasi awal. Selanjutnya dilakukan hyperparameter tuning terhadap contamination, n_estimators, dan max_samples. Parameter contamination merepresentasikan proporsi data yang diasumsikan sebagai anomali, n_estimators menentukan jumlah Isolation Tree, dan max_samples mengatur jumlah data yang digunakan pada setiap pohon. Setelah konfigurasi hyperparameter dipilih, anomaly score digunakan pada threshold tuning untuk menentukan batas keputusan yang memberikan keseimbangan terbaik antara Precision dan Recall berdasarkan F1-Score.

Model akhir dievaluasi menggunakan Confusion Matrix, Accuracy, Precision, Recall, F1-Score, ROC-AUC, dan PR-AUC. Analisis juga mencakup False Positive, False Negative, serta distribusi anomaly score. Setelah itu, Isolation Forest dibandingkan dengan LOF menggunakan dataset dan preprocessing yang sama agar perbandingan dilakukan pada kondisi yang konsisten.

4. HASIL DAN PEMBAHASAN

Karakteristik Data dan Hasil Preprocessing

Exploratory Data Analysis menunjukkan beberapa pola yang relevan. Debit card, perangkat mobile, dan kategori Toys & Games memiliki persentase fraud tertinggi sekitar 5,05%, tetapi selisih antar kategori sangat kecil. Transaksi fraud cenderung berasal dari akun yang lebih baru dengan median sekitar 63 hari, memiliki nominal transaksi yang lebih tinggi daripada transaksi normal, dan lebih banyak terjadi pada pukul 00.00–05.00. Customer Age dan Quantity tidak menunjukkan perbedaan yang signifikan. Temuan ini digunakan sebagai konteks analisis, bukan sebagai aturan tunggal penentuan fraud.

Tabel 6. Ringkasan Hasil EDA

Atribut	Temuan
Payment Method	Debit card memiliki persentase fraud tertinggi (5,05%), tetapi perbedaannya kecil.
Device Used	Mobile memiliki persentase fraud tertinggi (5,05%), selisih tidak signifikan.
Product Category	Toys & Games memiliki persentase fraud tertinggi (5,05%), tidak jauh berbeda dari kategori lain.
Account Age Days	Fraud cenderung berasal dari akun lebih baru; median sekitar 63 hari.
Transaction Amount	Transaksi fraud memiliki nominal lebih tinggi dibanding transaksi normal.
Transaction Hour	Fraud lebih banyak terjadi pukul 00.00–05.00.
Customer Age	Tidak terdapat perbedaan yang signifikan.
Quantity	Tidak terdapat perbedaan yang signifikan.

Pengaruh Preprocessing

Hasil eksperimen menunjukkan perbedaan besar antara kedua skenario preprocessing. Skenario A menghasilkan Recall tinggi tetapi Precision rendah, yang menunjukkan model menemukan banyak transaksi fraud namun juga menghasilkan banyak false positive. Skenario B menurunkan Recall sedikit, tetapi meningkatkan Precision secara substansial sehingga F1-Score dan Accuracy menjadi lebih tinggi.

Tabel 7. Pengaruh Preprocessing terhadap Kinerja Model

Skenario	Precision	Recall	F1-Score	Accuracy
Skenario A	0,2214	0,8618	0,3522	0,8412
Skenario B	0,8201	0,8200	0,8200	0,9820

Skenario B dipilih sebagai konfigurasi preprocessing model akhir karena memberikan keseimbangan yang lebih baik antara Precision dan Recall. Namun, peningkatan ini harus diinterpretasikan sebagai pengaruh kombinasi feature engineering dan StandardScaler karena keduanya tidak diuji secara terpisah.

Baseline, Hyperparameter, dan Threshold Tuning

Model baseline Isolation Forest menghasilkan Recall yang relatif tinggi, tetapi Precision dan F1-Score masih rendah. Setelah rangkaian preprocessing dan tuning, model akhir memperlihatkan peningkatan besar terutama pada Precision, F1-Score, dan Accuracy. Perbandingan baseline dan model akhir digunakan untuk menunjukkan perubahan kualitas model selama eksperimen.

Tabel 8. Perbandingan Baseline dan Model Akhir

Model	Precision	Recall	F1-Score	Accuracy
Baseline IF	0,2301	0,8571	0,3628	0,8492
Model Akhir IF	0,8201	0,8200	0,8200	0,9820

Pengujian contamination menggunakan nilai 0,01; 0,03; 0,05; 0,07; 0,10; 0,15; dan 0,20. Nilai 0,05 memberikan F1-Score tertinggi pada tahap ini, yaitu 0,4136, sehingga digunakan pada pengujian berikutnya.

Tabel 9. Pengaruh Contamination terhadap Kinerja Model

Contamination	Precision	Recall	F1-Score	Accuracy
0,01	0,2350	0,8567	0,3688	0,8531
0,03	0,2384	0,8735	0,3746	0,8539
0,05	0,2723	0,8596	0,4136	0,8779
0,07	0,2648	0,8482	0,4037	0,8745
0,10	0,2489	0,8597	0,3680	0,8630
0,15	0,2624	0,8697	0,4032	0,8710
0,20	0,2435	0,8660	0,3801	0,8558

Pengujian jumlah pohon menunjukkan bahwa n_estimators = 250 menghasilkan F1-Score tertinggi sebesar 0,4030 pada tahap tersebut. Penambahan jumlah pohon cenderung meningkatkan kestabilan proses isolasi, tetapi peningkatannya tidak selalu linear.

Tabel 10. Pengaruh n_estimators terhadap Kinerja Model

n_estimators	Precision	Recall	F1-Score	Accuracy
50	0,2361	0,8510	0,3696	0,8546
100	0,2257	0,8592	0,3575	0,8453
150	0,2587	0,8747	0,3993	0,8682
200	0,2587	0,8793	0,3993	0,8678
250	0,2618	0,8754	0,4030	0,8701

Pada pengujian max_samples, nilai 0,80 menghasilkan F1-Score tertinggi sebesar 0,4097. Variasi max_samples memperlihatkan bahwa rasio sampel yang terlalu kecil atau terlalu besar belum tentu meningkatkan performa model.

Tabel 11. Pengaruh max_samples terhadap Kinerja Model

max_samples	Precision	Recall	F1-Score	Accuracy
auto	0,2368	0,8725	0,3726	0,8528
0,5	0,2596	0,8495	0,3977	0,8711
0,6	0,2498	0,8462	0,3857	0,8650
0,7	0,2328	0,8495	0,3654	0,8522
0,8	0,2687	0,8617	0,4097	0,8756
0,9	0,2256	0,8601	0,3574	0,8451
1,0	0,2608	0,8436	0,3985	0,8724

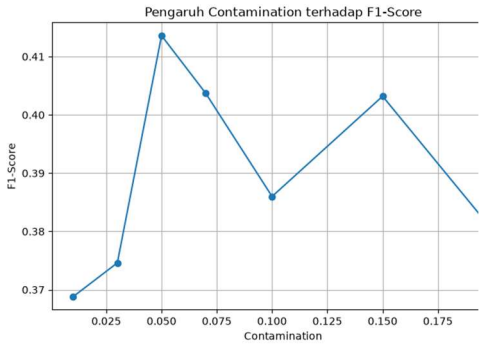
Threshold tuning dilakukan setelah hyperparameter dipilih. Threshold 0,45 menghasilkan F1-Score tertinggi 0,4157 pada pengujian ambang batas dan selanjutnya digunakan bersama contamination 0,05, n_estimators 250, dan max_samples 0,80.

Tabel 12. Ringkasan Threshold Tuning

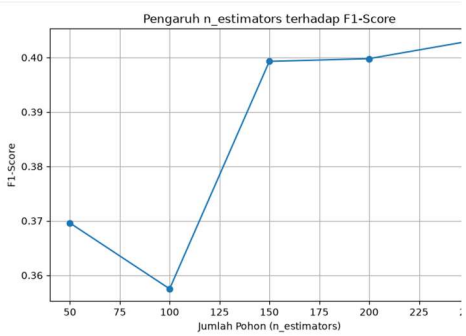
Threshold	Accuracy	Precision	Recall	F1-Score
0,30	0,8635	0,2511	0,8702	0,3898
0,40	0,8777	0,2723	0,8621	0,4139
0,45	0,8771	0,2728	0,8729	0,4157
0,70	0,8771	0,2689	0,8461	0,4081
0,80	0,8740	0,2676	0,8730	0,4096

Tabel 13. Konfigurasi Model Akhir Isolation Forest

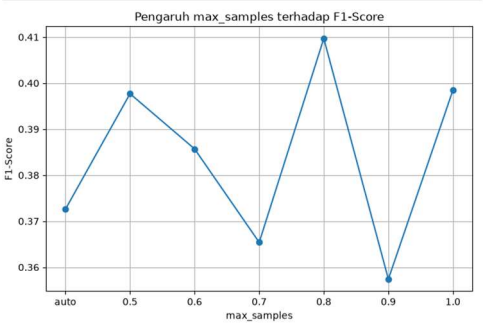
Parameter	Nilai
contamination	0,05
n_estimators	250
max_samples	0,80
threshold	0,45
random_state	42



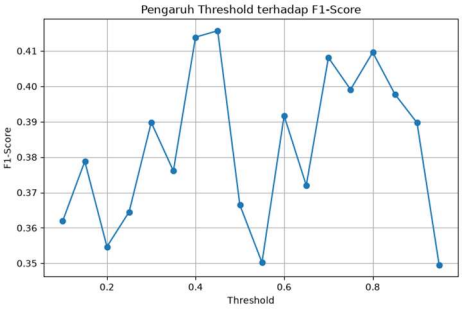
(a) Contamination



(b) n_estimators



(c) max_samples

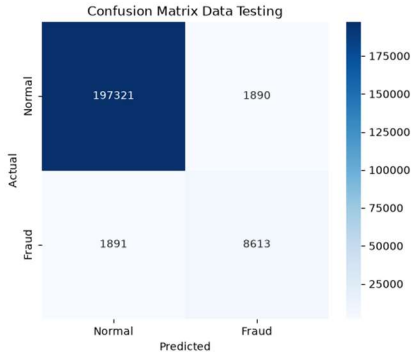


(d) Threshold

Gambar 2. Pengaruh Parameter dan Threshold terhadap F1-Score

Evaluasi Model Isolation Forest

Dengan konfigurasi akhir, Isolation Forest menghasilkan Accuracy 0,9820, Precision 0,8201, Recall 0,8200, dan F1-Score 0,8200. Precision 0,8201 menunjukkan bahwa sekitar 82,01% transaksi yang diprediksi fraud merupakan fraud aktual, sedangkan Recall 0,8200 menunjukkan bahwa sekitar 82,00% transaksi fraud berhasil ditemukan. Keseimbangan kedua metrik tercermin pada F1-Score sebesar 0,8200.



Gambar 3. Confusion Matrix Isolation Forest

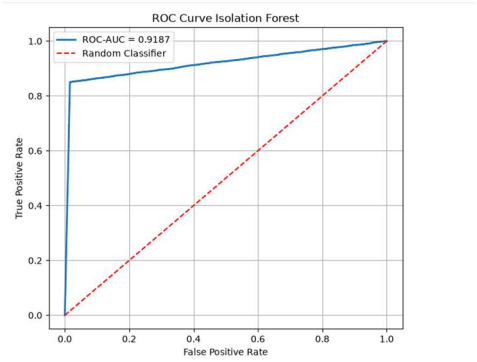
Tabel 14. Confusion Matrix Isolation Forest

Aktual	Prediksi Normal	Prediksi Fraud
Normal	197.321	1.890
Fraud	1.891	8.613

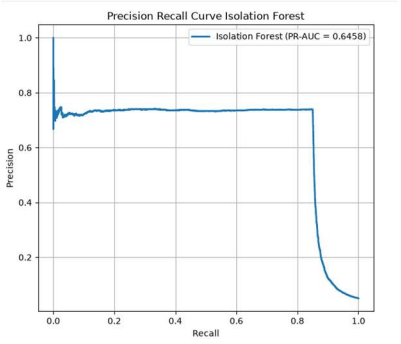
Confusion Matrix menunjukkan 197.321 True Negative dan 8.613 True Positive. Kesalahan terdiri atas 1.890 False Positive dan 1.891 False Negative. Jumlah False Negative tetap penting karena merepresentasikan transaksi fraud yang gagal terdeteksi, sementara False Positive dapat menyebabkan transaksi normal ikut ditandai sebagai mencurigakan.

ROC-AUC, PR-AUC, dan Anomaly Score

Model memperoleh ROC-AUC sebesar 0,9187 dan PR-AUC sebesar 0,6458. Berdasarkan interpretasi yang digunakan pada skripsi, ROC-AUC 0,9187 menunjukkan kemampuan diskriminasi yang tinggi antara transaksi fraud dan normal. PR-AUC digunakan bersama ROC-AUC karena dataset bersifat imbalanced dan transaksi fraud merupakan kelas minoritas.



(a) Kurva ROC Isolation Forest



(b) Kurva Precision-Recall Isolation Forest

Gambar 4. ROC dan Precision-Recall Isolation Forest

Distribusi anomaly score memperlihatkan mayoritas transaksi memiliki skor rendah, sedangkan hanya sebagian kecil transaksi berada pada skor tinggi. Skor yang lebih tinggi menandakan penyimpangan yang lebih kuat dari pola mayoritas dan digunakan sebagai dasar threshold untuk klasifikasi anomali.

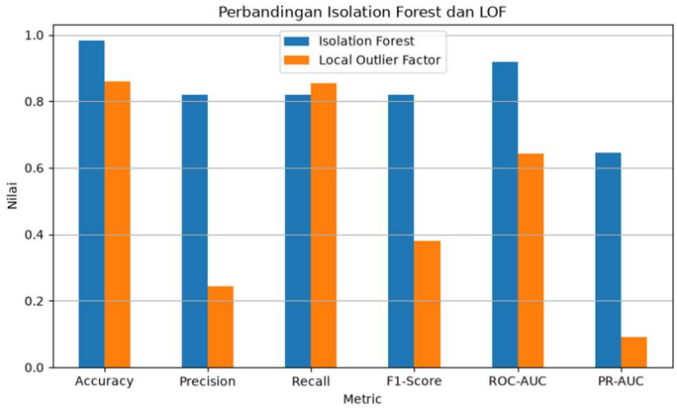
Perbandingan Isolation Forest dan LOF

LOF menghasilkan Precision 0,2451, Recall 0,8533, F1-Score 0,3809, Accuracy 0,8610, ROC-AUC 0,6417, dan PR-AUC 0,0916. Recall LOF sedikit lebih tinggi daripada Isolation Forest, tetapi Precision yang rendah menunjukkan banyak transaksi

normal dikategorikan sebagai fraud. Hal ini juga terlihat dari False Positive LOF yang mencapai 27.600 transaksi, dibandingkan 1.890 pada Isolation Forest.

Tabel 15. Perbandingan Kinerja Isolation Forest dan LOF

Metode	Precision	Recall	F1-Score	ROC-AUC	PR-AUC	Accuracy
Isolation Forest	0,8201	0,8200	0,8200	0,9187	0,6458	0,9820
LOF	0,2451	0,8533	0,3809	0,6417	0,0916	0,8610



Gambar 5. Perbandingan Kinerja Isolation Forest dan LOF

Secara keseluruhan, Isolation Forest memberikan performa lebih tinggi pada Precision, F1-Score, ROC-AUC, PR-AUC, dan Accuracy, sedangkan LOF unggul sedikit pada Recall. Hasil ini menunjukkan pentingnya mempertimbangkan beberapa metrik secara bersamaan. Recall tinggi bermanfaat untuk menemukan fraud, tetapi Precision rendah meningkatkan beban pemeriksaan karena banyak transaksi normal ikut terdeteksi sebagai fraud.

Pembahasan

Hasil penelitian memperlihatkan bahwa Isolation Forest dapat digunakan untuk deteksi anomali pada transaksi marketplace yang sangat tidak seimbang. Konfigurasi akhir contamination 0,05, n_estimators 250, max_samples 0,80, threshold 0,45, dan random_state 42 menghasilkan keseimbangan Precision dan Recall yang baik serta ROC-AUC tinggi. Temuan ini konsisten dengan penelitian yang menunjukkan efektivitas Isolation Forest pada data fraud yang imbalanced [11], [14].

Perbandingan preprocessing menunjukkan bahwa transformasi data tidak hanya berfungsi sebagai tahap persiapan, tetapi dapat memengaruhi hasil deteksi. Pada penelitian ini, kombinasi feature engineering dan StandardScaler dalam Skenario B

meningkatkan Precision, F1-Score, dan Accuracy secara besar dibandingkan Skenario A. Meskipun Isolation Forest tidak bergantung langsung pada skala atribut seperti algoritma berbasis jarak, StandardScaler digunakan untuk konsistensi representasi data dan perbandingan yang lebih adil dengan LOF.

Hasil deteksi tidak ditujukan sebagai keputusan mutlak bahwa transaksi tertentu merupakan fraud. Model berperan sebagai mekanisme penyaringan awal untuk mengidentifikasi transaksi yang memiliki pola menyimpang. Dalam penerapan nyata, prediksi sebaiknya dipadukan dengan verifikasi tambahan, riwayat transaksi, reputasi akun, lokasi, dan mekanisme autentikasi agar False Positive dapat ditekan.

Keterbatasan dan Implikasi Penelitian

Penelitian menggunakan dataset publik Fraudulent E-Commerce Transactions dan tidak melakukan pengumpulan data langsung dari platform marketplace tertentu. Karena itu, hasil eksperimen merepresentasikan kemampuan model pada dataset yang digunakan, bukan validasi terhadap seluruh kondisi operasional marketplace. Faktor di luar atribut dataset, seperti kebijakan platform, perilaku pengguna di luar transaksi, serta informasi nonteknis tidak dianalisis.

Keterbatasan lain terdapat pada eksperimen preprocessing. Feature engineering dan StandardScaler diterapkan secara bersamaan pada Skenario B, sehingga peningkatan performa tidak dapat diatribusikan secara terpisah pada salah satu tahap. Penelitian juga tidak menggunakan resampling maupun pendekatan hybrid, karena fokus utama diarahkan pada evaluasi Isolation Forest dan sensitivitas konfigurasi model.

Implikasi praktis dari hasil penelitian adalah penggunaan Isolation Forest sebagai alat penyaringan awal transaksi mencurigakan. Model dapat membantu memprioritaskan transaksi yang memerlukan pemeriksaan lebih lanjut, tetapi keputusan akhir tetap perlu mempertimbangkan verifikasi tambahan dan informasi kontekstual. Pendekatan ini penting untuk mengurangi risiko keputusan otomatis yang salah, terutama pada kasus False Positive ketika transaksi normal memiliki karakteristik tidak umum.

Tabel 16. Keterbatasan Utama dan Arah Pengembangan

Aspek	Keterbatasan / Arah Pengembangan
Sumber data	Menggunakan dataset publik; penelitian berikutnya dapat menguji data transaksi aktual jika akses dan privasi memungkinkan.
Preprocessing	Feature engineering dan StandardScaler belum diuji secara terpisah.

Aspek	Keterbatasan / Arah Pengembangan
Optimasi parameter	Dapat dikembangkan dengan Grid Search, Random Search, atau Bayesian Optimization.
Metode pembanding	Dapat diperluas ke One-Class SVM, DBSCAN, Autoencoder, dan deep learning.
Implementasi	Hasil model perlu dikombinasikan dengan verifikasi kontekstual sebelum digunakan dalam keputusan operasional.

5. KESIMPULAN DAN SARAN

Penelitian menunjukkan bahwa preprocessing berpengaruh terhadap performa Isolation Forest. Skenario A menghasilkan Precision 0,2214, Recall 0,8618, F1-Score 0,3522, dan Accuracy 0,8412, sedangkan Skenario B menghasilkan Precision 0,8201, Recall 0,8200, F1-Score 0,8200, dan Accuracy 0,9820. Karena feature engineering dan StandardScaler diuji secara bersamaan, peningkatan performa dipahami sebagai pengaruh kombinasi kedua tahap tersebut.

Konfigurasi akhir model terdiri atas contamination 0,05, n_estimators 250, max_samples 0,80, threshold 0,45, dan random_state 42. Model memperoleh ROC-AUC 0,9187 dan PR-AUC 0,6458 serta Confusion Matrix dengan 197.321 True Negative, 8.613 True Positive, 1.890 False Positive, dan 1.891 False Negative. Dibandingkan LOF, Isolation Forest menghasilkan keseimbangan Precision–Recall yang lebih baik dan unggul pada sebagian besar metrik evaluasi.

Untuk penelitian berikutnya, disarankan menggunakan dataset yang lebih beragam atau data transaksi aktual jika akses dan privasi memungkinkan; menguji feature engineering dan StandardScaler secara terpisah; menggunakan pencarian parameter yang lebih sistematis seperti Grid Search, Random Search, atau Bayesian Optimization; serta membandingkan Isolation Forest dengan One-Class SVM, DBSCAN, Autoencoder, atau pendekatan deep learning. Pengembangan menuju sistem penyaringan awal fraud perlu dikombinasikan dengan verifikasi kontekstual agar keputusan operasional tidak hanya bergantung pada keluaran model.

DAFTAR REFERENSI

[1] M. Sánchez-Paniagua, E. Fidalgo, E. Alegre, and F. Jáñez-Martino, “Fraud detection in e-commerce: a comparative analysis of features to enhance machine learning models,” Electronic Commerce Research, vol. 26, no. 2, pp. 2467–2502, Apr. 2026, doi: 10.1007/s10660-025-10029-9.

[2] S. R. Byrapu Reddy, P. Kanagala, P. Ravichandran, D. R. Pulimamidi, P. V. Sivarambabu, and N. S. A. Polireddi, “Effective fraud detection in e-commerce:

Leveraging machine learning and big data analytics," Measurement: Sensors, vol. 33, Jun. 2024, doi: 10.1016/j.measen.2024.101138.

[3] L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, and M. S. Rodríguez Barrero, "Financial fraud detection through the application of machine learning techniques: a literature review," Springer Nature, Dec. 2024, doi: 10.1057/s41599-024-03606-0.

[4] S. Renjith, "Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine Approach," International Journal of Engineering Trends and Technology (IJETT), vol. 57, no. 1, pp. 48–53, Mar. 2018.

[5] S. Gavrylenko and O. Abdullin, "Improving the Accuracy of Machine Learning Model for Fraud Transactions with Combined Methods for Transaction Analysis," Innovative Technologies and Scientific Solutions for Industries, vol. 2024, no. 4, pp. 31–38, Dec. 2024, doi: 10.30837/2522-9818.2024.4.031.

[6] Y. F. Zhang, H. L. Lu, H. F. Lin, X. C. Qiao, and H. Zheng, "The Optimized Anomaly Detection Models Based on an Approach of Dealing with Imbalanced Dataset for Credit Card Fraud Detection," Mobile Information Systems, vol. 2022, 2022, doi: 10.1155/2022/8027903.

[7] I. Waspada, N. Bahtiar, P. W. Wirawan, B. Dwi, and A. Awan, "Performance Analysis of Isolation Forest Algorithm in Fraud Detection of Credit Card Transactions," 2020.

[8] F. Leveni, G. Weigert Cassales, B. Pfahringer, A. Bifet, and G. Boracchi, "Online Isolation Forest," in Proceedings of the 41st International Conference on Machine Learning, vol. 235, pp. 27288–27298, 2024.

[9] S. Ounacer, H. Ait, E. Bour, Y. Oubrahim, M. Y. Ghoumari, and M. Azzouazi, "Using Isolation Forest in anomaly detection: the case of credit card transactions," vol. 6, no. 2, pp. 394–400, 2018.

[10] Y. F. Zhang, H. L. Lu, H. F. Lin, X. C. Qiao, and H. Zheng, "The Optimized Anomaly Detection Models Based on an Approach of Dealing with Imbalanced Dataset for Credit Card Fraud Detection," Mobile Information Systems, vol. 2022, 2022, doi: 10.1155/2022/8027903.

[11] I. Waspada, N. Bahtiar, P. W. Wirawan, B. Dwi, and A. Awan, "Performance Analysis of Isolation Forest Algorithm in Fraud Detection of Credit Card Transactions," 2020.

[12] R. S. Ismanda, M. T. A. Silitonga, and S. Nur Hasanah, "Deteksi Hybrid Anomali Transaksi Digital dengan Optimasi Isolation Forest-K-Means untuk Peningkatan Keamanan Finansial," Innovative: Journal of Social Science Research, vol. 5, no. 3, pp. 5749–5765, 2025, doi: 10.31004/innovative.v5i3.19791.

[13] U. Porwal and S. Mukund, "Credit Card Fraud Detection in e-Commerce: An Outlier Detection Approach," International Journal of Computer Sciences and Engineering, May 2019.

[14] H. John and S. Naaz, "Credit Card Fraud Detection using Local Outlier Factor and Isolation Forest," International Journal of Computer Sciences and Engineering, vol. 7, no. 4, pp. 1060–1064, Apr. 2019, doi: 10.26438/ijcse/v7i4.10601064.

[15] A. Alsarhan, F. Hussein, S. Moh, and F. S. El-Salhi, "The Effect of Preprocessing Techniques, Applied to Numeric Features, on Classification Algorithms' Performance," 2021, doi: 10.3390/data.

[16] A. Alsarhan, F. Hussein, S. Moh, and F. S. El-Salhi, "The Effect of Preprocessing Techniques, Applied to Numeric Features, on Classification Algorithms' Performance," Data (Basel), 2021, doi: 10.3390/data.
[17] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[18] L. Becker, "Anomaly Detection in E-Commerce Platforms via Graph Neural Networks," Asia Pacific Economic and Management Review, vol. 2, no. 2, Mar. 2025, doi: 10.62177/apemr.v2i2.208.
[19] S. Ounacer, H. Ait, E. Bour, Y. Oubrahim, M. Y. Ghoumari, and M. Azzouazi, "Using Isolation Forest in anomaly detection: the case of credit card transactions," J. Big Data, vol. 6, no. 2, pp. 394–400, 2018.
[20] N. Goix, "How to Evaluate the Quality of Unsupervised Anomaly Detection Algorithms?," Jul. 2016.
[21] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, no. 1, Jan. 2020, doi: 10.1186/s12864-019-6413-7.
[22] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed. Hoboken, NJ, USA: John Wiley & Sons, 2013, doi: 10.1002/9781118548387.
[23] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," PLoS One, vol. 10, no. 3, Mar. 2015, doi: 10.1371/journal.pone.0118432.
[24] Y. F. Zhang, H. L. Lu, H. F. Lin, X. C. Qiao, and H. Zheng, "The Optimized Anomaly Detection Models Based on an Approach of Dealing with Imbalanced Dataset for Credit Card Fraud Detection," Mobile Information Systems, vol. 2022, 2022, doi: 10.1155/2022/8027903.
[25] S. Gavrylenko and O. Abdullin, "Improving the Accuracy of Machine Learning Model for Fraud Transactions with Combined Methods for Transaction Analysis," Innovative Technologies and Scientific Solutions for Industries, vol. 2024, no. 4, pp. 31–38, Dec. 2024, doi: 10.30837/2522-9818.2024.4.031.
[26] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in Proc. 8th IEEE Int. Conf. Data Mining (ICDM), 2008, pp. 413–422, doi: 10.1109/ICDM.2008.17.