BETRIK
BESEMAH TEKNOLOGI INFORMASI & KOMPUTER
Editor Office : Pusat Penelitian & Pengabdian Pada Masyarakat (PPPM) ITPA
Phone : 0857-9716-9578
email : betrikitpa@itpa.ac.id
ISSN : 2339 - 1871

Perbandingan Kinerja *Isolation Forest* Dan *Local Outlier Factor* (LOF) Dalam Deteksi Anomali Transaksi Digital

Sri Hartati[1], Defi Pujianto[2], Kadarsih[3]
Fakultas STTM, Prodi D3 Manajemen Informatika, Universitas Mahakarya Asia, OKU,Indonesia[1,2]
Fakultas STTM, Prodi D3 Teknik Informatika, Universitas Mahakarya Asia, OKU,Indonesia[3]
Sur-el :*hartatiakmi1984@gmail.com[1], dhelphie85 @gmail.com[2] , kadarsih.mail@gmail.com[3]

Penulis Korespondensi: Sri Hartati, hartatiakmi1984@gmail.com

Abstrak: Perkembangan transaksi digital yang pesat meningkatkan risiko terjadinya aktivitas anomali seperti *fraud*, terutama pada dataset yang sangat tidak seimbang, di mana jumlah transaksi *fraud* jauh lebih sedikit dibandingkan transaksi normal. Ketidakseimbangan data menyebabkan model pembelajaran mesin cenderung mengutamakan pola dari kelas mayoritas, sehingga mengurangi kemampuan dalam mengenali anomali yang jumlahnya terbatas. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Isolation Forest* dan *Local Outlier Factor* (LOF) dalam mendeteksi anomali pada data transaksi digital. Metode penelitian menggunakan pendekatan eksperimen dengan dataset *Credit Card Fraud Detection* yang terdiri dari 284.807 transaksi, termasuk 492 transaksi fraud. Prapemrosesan data dilakukan melalui normalisasi fitur menggunakan *StandardScaler,* kemudian data dibagi menjadi data latih dan data uji dengan rasio 70:30 secara *stratified*. Evaluasi model dilakukan menggunakan confusion matrix, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa *Isolation Forest* memiliki performa yang lebih baik dibandingkan LOF. *Isolation Forest* mampu mendeteksi 37 dari 148 transaksi fraud dengan nilai precision sebesar 0,2824, recall sebesar 0,25, dan F1-*score* sebesar 0,2652. Sementara itu, LOF hanya mampu mendeteksi 2 transaksi f*raud* dengan *precision* sebesar 0,0137, *recall* sebesar 0,0135, dan F1-score sebesar 0,0136. Temuan ini menunjukkan bahwa pendekatan berbasis isolasi lebih efektif dan stabil dibandingkan pendekatan berbasis kepadatan dalam menangani dataset yang sangat tidak seimbang.

Kata kunci : deteksi anomali*, isolation forest, local outlier factor, fraud detection, dataset* tidak seimbang*,* transaksi digital*, unsupervised learning*

Abstract : *The rapid growth of digital transactions has increased the risk of anomalous activities such as fraud, particularly in highly imbalanced datasets where fraudulent transactions are significantly fewer than normal transactions. This imbalance presents a major challenge in anomaly detection, as models tend to be biased toward the majority class. This study aims to compare the performance of Isolation Forest and Local Outlier Factor (LOF) algorithms in detecting anomalies in digital transaction data.The research adopts an experimental approach using the Credit Card Fraud Detection dataset, which consists of 284,807 transactions, including 492 fraudulent cases. Data preprocessing involves feature normalization using StandardScaler, followed by a stratified train-test split with a ratio of 70:30. Model evaluation is conducted using confusion matrix, precision, recall, and F1-score metrics.The results show that Isolation Forest outperforms LOF. Isolation Forest successfully detects 37 out of 148 fraudulent transactions with a precision of 0.2824, recall of 0.25, and F1-score of 0.2652. In contrast, LOF detects only 2 fraudulent transactions,*

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

with a precision of 0.0137, recall of 0.0135, and F1-score of 0.0136. These findings indicate that isolation-based approaches are more effective and robust than density-based methods in handling highly imbalanced datasets.

Keywords: *anomaly detection, isolation forest, local outlier factor, fraud detection, imbalanced dataset, digital transactions, unsupervised learning*

1. PENDAHULUAN

Perkembangan sistem pembayaran digital dalam beberapa tahun terakhir telah mendorong peningkatan volume transaksi elektronik secara signifikan di berbagai sektor keuangan, termasuk perbankan, e-commerce, dan layanan fintech yang semakin terintegrasi dalam kehidupan masyarakat modern. Transformasi digital ini memberikan kemudahan dalam bertransaksi, namun di sisi lain juga membuka peluang terjadinya berbagai bentuk kejahatan siber, khususnya aktivitas kecurangan (*fraud*) yang semakin kompleks dan sulit dideteksi[1].

Peningkatan aktivitas transaksi digital yang masif tersebut menyebabkan sistem keuangan menghadapi tantangan baru dalam menjaga keamanan dan integritas data transaksi. Fraud dalam transaksi digital tidak hanya berdampak pada kerugian finansial bagi institusi keuangan, tetapi juga dapat menurunkan tingkat kepercayaan pengguna terhadap sistem pembayaran digital yang digunakan[2]. Oleh karena itu, diperlukan sistem deteksi kecurangan yang mampu bekerja secara cepat, akurat, dan adaptif terhadap perubahan pola transaksi yang dinamis[3].

Salah satu permasalahan utama dalam deteksi fraud adalah karakteristik dataset transaksi yang sangat tidak seimbang (*imbalanced dataset*), di mana jumlah transaksi normal jauh lebih besar dibandingkan transaksi anomali. Kondisi ini menyebabkan model pembelajaran mesin cenderung bias terhadap kelas mayoritas sehingga mengabaikan pola penting yang terdapat pada kelas minoritas [4]. Ketidakseimbangan data ini juga berdampak pada rendahnya kemampuan model dalam mendeteksi transaksi fraud yang sebenarnya memiliki tingkat risiko tinggi terhadap kerugian finansial [5].

Dalam mengatasi permasalahan tersebut, berbagai pendekatan berbasis *machine learning* telah dikembangkan untuk meningkatkan efektivitas sistem deteksi fraud. Metode pembelajaran mesin mampu menganalisis pola transaksi berdasarkan data historis dan mengidentifikasi perilaku yang menyimpang dari pola normal dengan tingkat akurasi yang lebih tinggi dibandingkan metode konvensional [6]. Selain itu, model berbasis *machine learning* juga memiliki kemampuan untuk beradaptasi terhadap perubahan pola transaksi secara dinamis seiring dengan perkembangan sistem digital [7].

Seiring dengan meningkatnya kompleksitas data transaksi, pendekatan berbasis *unsupervised learning* menjadi semakin relevan dalam deteksi anomali. Pendekatan ini tidak memerlukan data berlabel secara lengkap dan mampu mengidentifikasi pola penyimpangan dengan mempelajari distribusi data normal yang tersedia [8]. Hal ini sangat penting dalam kasus fraud detection, karena data transaksi yang teridentifikasi sebagai fraud umumnya sangat terbatas dan sulit diperoleh dalam jumlah besar [9].

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

Di antara berbagai algoritma deteksi anomali yang tersedia, Isolation Forest dan Local Outlier Factor (LOF) merupakan dua metode yang banyak digunakan karena efisiensi dan kemampuannya dalam mendeteksi outlier. Isolation Forest merupakan algoritma berbasis isolasi yang bekerja dengan cara memisahkan data melalui partisi acak, di mana data anomali cenderung lebih mudah diisolasi dibandingkan data normal karena memiliki karakteristik yang berbeda dari mayoritas data [10]. Selain itu, metode ini memiliki keunggulan dalam menangani data berdimensi tinggi dan tidak bergantung pada distribusi data, sehingga lebih robust pada dataset dengan karakteristik tidak seimbang [11]. Sementara itu, Local Outlier Factor (LOF) merupakan metode berbasis kepadatan yang mengidentifikasi anomali dengan membandingkan kepadatan lokal suatu titik data terhadap tetangganya, sehingga mampu mendeteksi penyimpangan berdasarkan struktur lokal data [12]. Pendekatan Local Outlier Factor (LOF) mengidentifikasi anomali melalui perbandingan kepadatan lokal antar data terhadap tetangganya, sehingga mampu mengidentifikasi outlier berdasarkan struktur lokal data. Pendekatan berbasis kepadatan ini efektif dalam mendeteksi anomali lokal, namun performanya sangat dipengaruhi oleh distribusi data dan parameter yang digunakan. Pendekatan ini efektif dalam mendeteksi anomali lokal, namun performanya sangat dipengaruhi oleh distribusi data serta parameter yang digunakan dalam proses perhitungan kepadatan.

Meskipun kedua algoritma tersebut banyak digunakan dalam berbagai penelitian, performanya sangat dipengaruhi oleh karakteristik dataset yang digunakan, khususnya pada kondisi data yang tidak seimbang. Beberapa penelitian menunjukkan bahwa metode berbasis kepadatan seperti LOF cenderung mengalami penurunan performa pada dataset dengan distribusi ekstrem, sedangkan metode berbasis isolasi lebih stabil dalam mendeteksi anomali pada data berdimensi tinggi [13].

Berdasarkan permasalahan tersebut, masih diperlukan kajian komparatif untuk mengevaluasi kinerja algoritma deteksi anomali yang memiliki efisiensi komputasi tinggi serta mampu bekerja secara optimal pada dataset transaksi digital yang tidak seimbang. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Isolation Forest dan Local Outlier Factor dalam mendeteksi anomali pada data transaksi digital menggunakan dataset credit card fraud detection.

Kebaruan penelitian ini terletak pada analisis komparatif dua algoritma *unsupervised learning* yang relatif ringan secara komputasi dengan fokus pada keseimbangan *precision* dan *recall* dalam mendeteksi transaksi anomali. Selain itu, penelitian ini menggunakan pendekatan eksperimen yang reprodusibel berbasis Python pada platform Google Colab sehingga dapat dengan mudah diimplementasikan pada sistem deteksi fraud di dunia nyata.

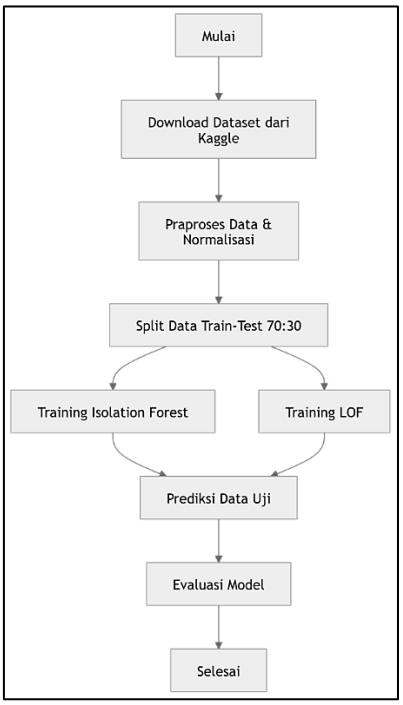
2. METODOLOGI PENELITIAN

Metode penelitian pada studi ini bertujuan untuk menguji dan membandingkan kinerja algoritma *Isolation Forest* dan *Local Outlier Factor* dalam mendeteksi anomali pada data transaksi digital. Proses penelitian dilakukan melalui beberapa tahapan, yaitu pengumpulan dataset, praproses data, pembangunan model deteksi anomali, serta evaluasi performa model menggunakan metrik evaluasi klasifikasi. Seluruh proses eksperimen dilakukan menggunakan bahasa pemrograman Python pada platform Google Colab.

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

Eksperimen dalam penelitian ini dilakukan dengan alur sebagai berikut:

1. Mengunduh dataset *credit card fraud detection* dari Kaggle
2. Melakukan praproses data dan normalisasi fitur
3. Membagi dataset menjadi data pelatihan dan data pengujian
4. Melatih model menggunakan algoritma Isolation Forest
5. Melatih model menggunakan algoritma Local Outlier Factor
6. Melakukan prediksi terhadap data pengujian
7. Mengevaluasi performa model menggunakan metrik evaluasi klasifikasi



Gambar 1. Alur penelitian

2.1 Pendekatan penelitian

Penelitian ini menggunakan pendekatan eksperimen komputasional untuk menganalisis dan membandingkan kinerja dua algoritma deteksi anomali, yaitu *Isolation Forest* dan *Local Outlier Factor* (LOF), dalam mendeteksi transaksi anomali pada data transaksi digital. Pendekatan eksperimen digunakan untuk mengevaluasi kemampuan masing-masing algoritma dalam mengidentifikasi transaksi fraud pada dataset yang memiliki karakteristik tidak seimbang (*imbalanced dataset*).

Metode deteksi anomali yang digunakan dalam penelitian ini termasuk dalam kategori *unsupervised anomaly detection*, yaitu pendekatan pembelajaran mesin yang mampu mengidentifikasi penyimpangan dari pola normal tanpa memerlukan data berlabel secara lengkap [20]. Pendekatan ini sangat relevan digunakan dalam sistem deteksi fraud karena pada praktiknya data transaksi yang terlabel fraud sering kali sangat terbatas dibandingkan data transaksi normal.

2.2 *Dataset* Penelitian

Dataset yang digunakan dalam penelitian ini adalah Credit Card Fraud Detection Dataset yang tersedia secara publik pada platform Kaggle. Dataset ini berisi transaksi kartu kredit yang dilakukan oleh

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

pemegang kartu di Eropa pada bulan September 2013. Dataset terdiri dari 284.807 transaksi, dengan 492 transaksi fraud, sehingga memiliki karakteristik dataset yang sangat tidak seimbang (*class imbalance*). Dataset memiliki 30 atribut, yang terdiri dari:

1. Time : waktu transaksi
2. Amount : nilai transaksi
3. V1 – V28 : fitur hasil transformasi *Principal Component Analysis* (PCA)
4. Class : label transaksi (0 = normal, 1 = fraud)

Dataset ini banyak digunakan dalam penelitian deteksi fraud karena mencerminkan karakteristik data transaksi nyata dengan tingkat ketidakseimbangan yang tinggi.

2.3 Praproses Data

Sebelum dilakukan proses pemodelan, dataset terlebih dahulu melalui tahap praproses data untuk meningkatkan kualitas data yang digunakan dalam penelitian.

Tahapan praproses yang dilakukan meliputi:

1. Pemisahan fitur dan label

   Dataset dipisahkan menjadi dua bagian yaitu fitur (*features*) dan label (*target*). Fitur terdiri dari seluruh atribut kecuali atribut *Class*, sedangkan atribut *Class* digunakan sebagai label untuk mengevaluasi hasil prediksi model.

2. Normalisasi data

   Normalisasi dilakukan menggunakan metode StandardScaler untuk memastikan setiap fitur memiliki distribusi yang sebanding dan mengurangi pengaruh skala data terhadap proses pembelajaran model. Normalisasi data penting dilakukan dalam algoritma berbasis jarak seperti LOF karena perbedaan skala fitur dapat mempengaruhi hasil perhitungan jarak antar data [11].

3. Pembagian dataset

   Dataset dibagi menjadi dua bagian yaitu:

   a. Data pelatihan (training data) sebesar 70%
   b. Data pengujian (testing data) sebesar 30%

   Pembagian dataset dilakukan menggunakan metode *train-test split* untuk memastikan model dapat dievaluasi menggunakan data yang tidak digunakan pada proses pelatihan[14].

2.4 Model Deteksi Anomali

Penelitian ini menggunakan dua algoritma deteksi anomali yang banyak digunakan dalam analisis data, yaitu *Isolation Forest* dan *Local Outlier Factor*.

1. Isolation Forest

Isolation Forest merupakan algoritma *unsupervised anomaly detection* yang bekerja berdasarkan prinsip isolasi data melalui proses partisi acak. Algoritma ini membangun sekumpulan pohon keputusan (*isolation trees*) yang digunakan untuk mengisolasi setiap titik data dalam dataset. Data anomali cenderung memiliki jalur isolasi yang lebih pendek dibandingkan data normal karena memiliki nilai fitur yang berbeda secara signifikan dari mayoritas data[10].

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

Skor anomali pada Isolation Forest dihitung menggunakan persamaan berikut :

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

Penjelasan :

$s(x,n)$= skor anomali

$E(h(x))$= panjang jalur rata-rata

$c(n)$= faktor normalisasi berdasarkan ukuran sampel

Semakin tinggi nilai skor anomali, maka semakin besar kemungkinan suatu data dikategorikan sebagai anomali. Algoritma Isolation Forest memiliki kompleksitas waktu yang relatif rendah sehingga mampu menangani dataset berskala besar secara efisien.

2. *Local Outlier Factor*

Local Outlier Factor (LOF) merupakan algoritma deteksi anomali berbasis kepadatan (*density-based anomaly detection*). Algoritma ini mengidentifikasi anomali dengan membandingkan kepadatan lokal suatu titik data terhadap kepadatan tetangga terdekatnya.

Menurut Breunig et al. (2000, p. 95), suatu titik data dianggap sebagai *outlier* apabila kepadatan lokalnya secara signifikan lebih rendah dibandingkan kepadatan tetangga di sekitarnya.

Nilai LOF dihitung menggunakan persamaan berikut:

$$LOF_k(p) = \frac{\sum_{o \in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)}}{|N_k(p)|} \quad \text{.....(2)}$$

Penjelasan :

$lrd_k(p)$= kepadatan lokal titik data

$N_k(p)$= himpunan tetangga terdekat

k= jumlah tetangga terdekat

Nilai LOF yang lebih besar dari 1 menunjukkan bahwa suatu titik data memiliki kemungkinan sebagai anomali.

2.5 Rancangan Eksperimen

Seluruh eksperimen dalam penelitian ini dilakukan menggunakan bahasa pemrograman *Python* pada lingkungan komputasi berbasis *cloud*, yaitu *Google Colab*. Implementasi algoritma pembelajaran mesin dilakukan menggunakan pustaka *Scikit-learn* yang menyediakan fungsi untuk proses pemodelan, pelatihan, dan evaluasi model secara terstruktur. Penggunaan lingkungan komputasi berbasis cloud memungkinkan proses eksperimen dilakukan secara efisien tanpa memerlukan perangkat keras dengan spesifikasi tinggi.

Rancangan eksperimen dimulai dari proses pengunduhan dataset *Credit Card Fraud Detection* dari *platform Kaggle*, kemudian dilanjutkan dengan tahap praproses data yang meliputi pemisahan fitur dan label serta normalisasi fitur menggunakan metode *StandardScaler* untuk menyamakan skala data. Setelah tahap

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

praproses selesai, dataset dibagi menjadi data pelatihan dan data pengujian dengan rasio 70:30 menggunakan teknik stratified sampling untuk menjaga distribusi kelas tetap konsisten pada kedua data.

Selanjutnya, model deteksi anomali dibangun menggunakan algoritma *Isolation Forest* dan *Local Outlier Factor* (LOF). Kedua model dilatih menggunakan data pelatihan dan kemudian digunakan untuk melakukan prediksi terhadap data pengujian. Tahap akhir eksperimen dilakukan dengan mengevaluasi performa model menggunakan *confusion matrix, precision, recall*, dan F1-*score* untuk mengukur kemampuan model dalam mendeteksi transaksi anomali pada dataset yang sangat tidak seimbang. Hasil evaluasi kedua algoritma kemudian dibandingkan untuk menentukan metode yang memiliki performa terbaik dalam mendeteksi anomali transaksi digital.

2.6 Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan beberapa metrik evaluasi klasifikasi yang umum digunakan pada dataset tidak seimbang.

1. *Precision*

Precision mengukur proporsi prediksi anomali yang benar terhadap seluruh prediksi anomali.

$$Precision = \frac{TP}{TP+FP} \quad \text{.....(3)}$$

2. *Recall*

Recall mengukur kemampuan model dalam mendeteksi seluruh data anomali yang sebenarnya.

$$Recall = \frac{TP}{TP+FN} \quad \text{.....(4)}$$

3. F1-*Score*

F1-score merupakan rata-rata harmonik antara precision dan recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \text{.....(5)}$$

4. Area Under Curve (AUC-ROC)

AUC-ROC digunakan untuk mengevaluasi kemampuan model dalam membedakan antara kelas normal dan kelas anomali pada berbagai nilai ambang klasifikasi.

Hasil evaluasi dari kedua algoritma kemudian dibandingkan untuk menentukan algoritma yang memiliki performa terbaik dalam mendeteksi transaksi anomali pada dataset transaksi digital.

2.7 Hipotesis Penelitian

Hipotesis yang diajukan dalam penelitian ini adalah:

H1: Algoritma *Isolation Forest* memiliki performa deteksi anomali yang lebih baik dibandingkan algoritma *Local Outlier Factor* pada dataset transaksi digital yang memiliki karakteristik sangat tidak seimbang. Hipotesis ini didasarkan pada karakteristik Isolation Forest yang mampu mengisolasi anomali secara lebih efisien melalui proses partisi acak pada ruang fitur data

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Pembahasan hasil penelitian dilakukan berdasarkan tahapan eksperimen yang telah dirancang pada metodologi penelitian, mulai dari pengambilan dataset, praproses data, pembagian dataset, proses pelatihan model, hingga evaluasi performa algoritma dalam mendeteksi anomali transaksi digital. Bagian ini menyajikan hasil eksperimen yang diperoleh dari penerapan algoritma *Isolation Forest* dan *Local Outlier Factor* dalam mendeteksi anomali pada dataset transaksi digital. Hasil pengujian kemudian dianalisis untuk mengevaluasi kinerja masing-masing algoritma menggunakan metrik evaluasi seperti *precision*, *recall*, dan *F1-score*, sehingga dapat diketahui algoritma yang memiliki performa terbaik dalam mendeteksi transaksi anomali.

3.1.1 Analisis *Dataset*

Penelitian ini menggunakan *dataset Credit Card Fraud Detection* yang diperoleh dari platform *Kaggle*. Dataset terdiri dari 284.807 data transaksi dengan 30 atribut, yang meliputi fitur hasil transformasi Principal Component Analysis (PCA), atribut waktu transaksi (Time), nilai transaksi (Amount), serta label kelas transaksi (Class).

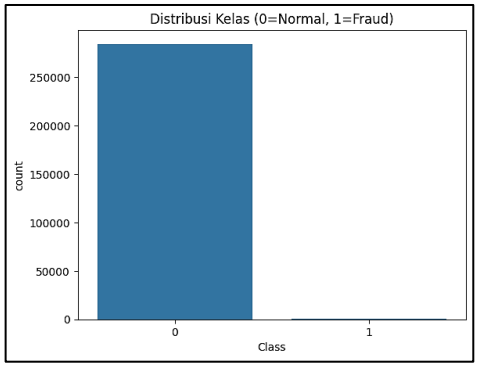
Ukuran dataset: (284807, 31)

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.128539	-0.189115	0.133558	-0.021053	149.62	0
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846	0.167170	0.125895	-0.008983	0.014724	2.69	0
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281	-0.327642	-0.139097	-0.055353	-0.059752	378.66	0
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175575	0.647376	-0.221929	0.062723	0.061458	123.50	0
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141267	-0.206010	0.502292	0.219422	0.215153	69.99	0

5 rows × 31 columns

Gambar 2 . Isi Dataset

Eksperimen dilakukan menggunakan *dataset Credit Card Fraud Detection*. Dataset ini memiliki karakteristik *imbalanced dataset* karena jumlah transaksi normal jauh lebih besar dibandingkan transaksi anomali. Distribusi data menunjukkan ketidakseimbangan yang sangat tinggi, di mana proporsi data fraud hanya sekitar 0,17% dari keseluruhan dataset. Kondisi ini menjadi tantangan utama dalam proses deteksi anomali karena model cenderung bias terhadap kelas mayoritas.



Gambar 3. Distribusi kelas

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

Dataset dibagi menggunakan metode *stratified train-test split* dengan perbandingan 70:30, sehingga diperoleh 199.364 data latih dan 85.443 data uji. Seluruh fitur kemudian dinormalisasi menggunakan StandardScaler untuk memastikan keseragaman skala data, terutama untuk mendukung kinerja algoritma berbasis jarak seperti LOF.

Pada penelitian ini dilakukan perbandingan kinerja antara algoritma *Isolation Forest* dan *Local Outlier Factor* (LOF) dalam mendeteksi anomali pada data transaksi digital. Kedua model dilatih menggunakan data latih dan diuji pada data uji untuk mengevaluasi kemampuan deteksi anomali.

3.1.2 Praproses Data

Tahap praproses data dilakukan sebelum proses pemodelan untuk meningkatkan kualitas data yang digunakan dalam penelitian. Pada tahap ini dilakukan pemisahan antara fitur dan label transaksi. Fitur terdiri dari seluruh atribut selain Class, sedangkan atribut Class digunakan sebagai label untuk membedakan transaksi normal dan transaksi fraud.

Selanjutnya dilakukan proses normalisasi fitur menggunakan metode StandardScaler. Normalisasi dilakukan untuk memastikan setiap fitur memiliki skala yang seragam sehingga dapat meningkatkan performa model, khususnya pada algoritma berbasis jarak seperti Local Outlier Factor (LOF).

Hasil normalisasi menunjukkan bahwa data berhasil ditransformasikan ke dalam distribusi yang lebih seimbang antar fitur sehingga proses pembelajaran model dapat dilakukan secara lebih optimal.

3.1.2.1 Pembagian *Dataset*

Setelah tahap praproses selesai, dataset dibagi menjadi data pelatihan dan data pengujian menggunakan metode stratified train-test split dengan rasio 70:30. Teknik stratified digunakan untuk menjaga proporsi distribusi kelas fraud dan non-fraud tetap konsisten pada data latih maupun data uji.

Hasil pembagian dataset menghasilkan:

a. Data pelatihan sebanyak 199.364 data
b. Data pengujian sebanyak 85.443 data

Pembagian data ini bertujuan agar model dapat mempelajari pola transaksi pada data pelatihan dan kemudian diuji menggunakan data yang belum pernah digunakan sebelumnya sehingga hasil evaluasi menjadi lebih objektif.

3.1.2.2 Hasil Deteksi Anomali

Pada tahap deteksi anomali, penelitian ini menggunakan dua algoritma, yaitu Isolation Forest dan Local Outlier Factor (LOF). Kedua model dilatih menggunakan data pelatihan dan kemudian digunakan untuk melakukan prediksi terhadap data pengujian. Hasil prediksi model kemudian dievaluasi menggunakan confusion matrix untuk mengetahui

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

kemampuan masing-masing algoritma dalam mengidentifikasi transaksi fraud dan transaksi normal.

Tabel 1. Hasil confusion matrix

Algoritma	*True Negative*	*False Positive*	*False Negative*	*True Positive*
Isolation Forest	85.201	94	111	37
Local Outlier Factor (LOF)	85.151	144	146	2

Berdasarkan *confusion matrix* tersebut, dapat diketahui bahwa Isolation Forest memiliki kemampuan yang lebih baik dalam mendeteksi data anomali dibandingkan LOF, yang terlihat dari jumlah True Positive yang lebih tinggi dan False Negative yang lebih rendah. Isolation Forest berhasil mendeteksi 37 transaksi fraud, sedangkan LOF hanya mampu mendeteksi 2 transaksi fraud.

3.1.2.3 Hasil perbandingan kinerja model

Evaluasi performa model dilakukan menggunakan metrik precision, recall, dan F1-score. Hasil evaluasi dari kedua algoritma ditunjukkan pada Tabel 2.

Tabel 2. Perbandingan Kinerja Model

Model	*Precision*	*Recall*	*F1-Score*
Isolation Forest	0.282443	0.250000	0.265233
Local Outlier Factor (LOF)	0.013699	0.013514	0.013605

3.1.2.4 Evaluasi model

Pada tahap deteksi anomali, penelitian ini menggunakan dua algoritma, yaitu *Isolation Forest* dan *Local Outlier Factor* (LOF). Kedua model dilatih menggunakan data pelatihan dan kemudian digunakan untuk melakukan prediksi terhadap data pengujian. Hasil prediksi model kemudian dievaluasi menggunakan confusion matrix untuk mengetahui kemampuan masing-masing algoritma dalam mengidentifikasi transaksi fraud dan transaksi normal.

Selanjutnya, hasil evaluasi menggunakan metrik precision, recall, dan F1-score disajikan pada gambar berikut:

```
=== Isolation Forest ===
[[85201    94]
 [  111    37]]
              precision    recall  f1-score   support

           0     0.9987    0.9989    0.9988     85295
           1     0.2824    0.2500    0.2652       148

    accuracy                         0.9976     85443
   macro avg     0.6406    0.6244    0.6320     85443
weighted avg     0.9975    0.9976    0.9975     85443
```

Gambar 4. Hasil Evaluasi Isolation Forest

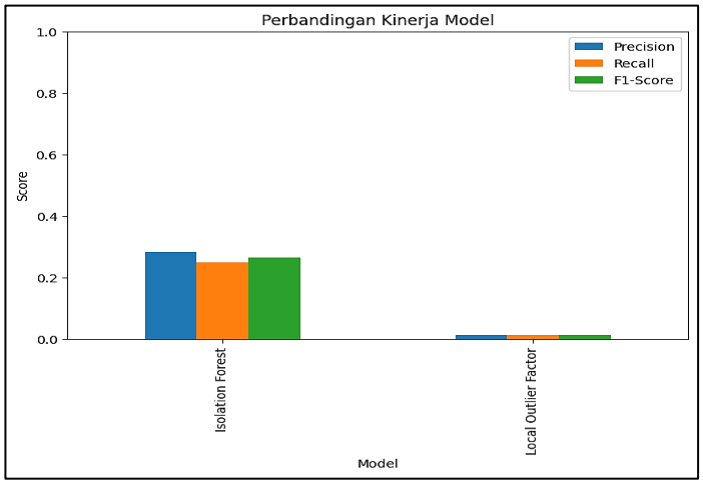
Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

```
=== Local Outlier Factor ===
[[85151   144]
 [  146     2]]
              precision    recall  f1-score   support

           0     0.9983    0.9983    0.9983     85295
           1     0.0137    0.0135    0.0136       148

    accuracy                         0.9966     85443
   macro avg     0.5060    0.5059    0.5060     85443
weighted avg     0.9966    0.9966    0.9966     85443
```

Gambar 5. Hasil Evaluasi LOF



Gambar 6. Visualisasi perbandingan model

3.2 Pembahasan

Berdasarkan hasil eksperimen yang telah dilakukan pada dataset Credit Card Fraud Detection, dapat diketahui bahwa karakteristik data yang digunakan dalam penelitian ini memiliki tingkat ketidakseimbangan yang sangat tinggi, di mana jumlah transaksi normal jauh lebih besar dibandingkan transaksi fraud. Kondisi ini menjadi tantangan utama dalam proses deteksi anomali karena model cenderung lebih mudah mengenali pola data normal dibandingkan data anomali yang jumlahnya sangat terbatas.

Hasil evaluasi menunjukkan bahwa kedua algoritma, yaitu Isolation Forest dan Local Outlier Factor (LOF), mampu melakukan klasifikasi terhadap data transaksi dengan tingkat akurasi yang tinggi. Namun demikian, nilai akurasi yang tinggi tidak sepenuhnya mencerminkan kemampuan model dalam mendeteksi anomali, karena dominasi kelas normal menyebabkan model tetap memperoleh nilai akurasi tinggi meskipun gagal mendeteksi sebagian besar transaksi fraud.

Jika ditinjau lebih lanjut menggunakan confusion matrix, terlihat perbedaan performa yang signifikan antara kedua algoritma. Isolation Forest mampu mendeteksi 37 dari 148 transaksi fraud pada data uji, sedangkan LOF hanya mampu mendeteksi 2 transaksi fraud. Selain itu, jumlah false negative pada LOF jauh lebih tinggi dibandingkan Isolation Forest, yang menunjukkan bahwa sebagian besar transaksi anomali

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

tidak berhasil teridentifikasi oleh LOF. Dalam konteks deteksi fraud, kondisi ini sangat tidak diharapkan karena dapat menyebabkan kerugian finansial akibat transaksi ilegal yang tidak terdeteksi.

Jika ditinjau dari metrik evaluasi, nilai recall pada Isolation Forest sebesar 0,25 menunjukkan bahwa model mampu mendeteksi 25% dari total anomali, sedangkan LOF hanya memiliki recall sebesar 0,0135. Hal ini menunjukkan bahwa LOF hampir tidak mampu mengidentifikasi anomali dalam dataset yang sangat tidak seimbang. Nilai precision dan F1-score yang jauh lebih tinggi pada Isolation Forest juga menunjukkan bahwa model ini memiliki keseimbangan performa yang lebih baik dalam mendeteksi anomali.

Perbedaan kinerja ini dipengaruhi oleh pendekatan dasar dari masing-masing algoritma. Isolation Forest bekerja dengan prinsip isolasi data melalui partisi acak, sehingga data yang memiliki karakteristik berbeda dari mayoritas akan lebih cepat terpisah dan teridentifikasi sebagai anomali. Pendekatan ini tidak bergantung pada distribusi kepadatan data, sehingga lebih robust terhadap dataset dengan distribusi ekstrem.

Sebaliknya, LOF merupakan metode berbasis kepadatan yang mengandalkan perbandingan kepadatan lokal antar titik data. Pada dataset dengan jumlah anomali yang sangat sedikit, kepadatan lokal data anomali sulit dibedakan dari data normal, sehingga menyebabkan rendahnya kemampuan deteksi. Hal ini menjelaskan mengapa LOF menghasilkan nilai recall dan F1-score yang sangat rendah dalam penelitian ini.

Selain itu, meskipun kedua model menunjukkan nilai akurasi yang tinggi, hasil penelitian ini menegaskan bahwa akurasi bukan merupakan metrik yang tepat dalam kasus dataset tidak seimbang. Model dengan akurasi tinggi belum tentu memiliki kemampuan yang baik dalam mendeteksi anomali. Oleh karena itu, metrik seperti recall dan F1-score menjadi indikator yang lebih relevan dalam mengevaluasi performa model deteksi fraud.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa Isolation Forest lebih efektif dan stabil dibandingkan LOF dalam mendeteksi anomali pada dataset transaksi digital yang sangat tidak seimbang. Temuan ini sejalan dengan hipotesis penelitian yang menyatakan bahwa pendekatan berbasis isolasi lebih unggul dibandingkan pendekatan berbasis kepadatan dalam kondisi distribusi data ekstrem. Dengan demikian, Isolation Forest dapat direkomendasikan sebagai metode yang lebih tepat untuk diimplementasikan dalam sistem deteksi fraud pada transaksi digital.

4. KESIMPULAN

4.1 Kesimpulan

Berdasarkan hasil eksperimen dan analisis yang telah dilakukan, penelitian ini menunjukkan bahwa karakteristik dataset transaksi digital yang sangat tidak seimbang memberikan tantangan signifikan dalam proses deteksi anomali. Proporsi data fraud yang sangat kecil menyebabkan model cenderung bias terhadap kelas normal, sehingga metrik akurasi tidak dapat dijadikan indikator utama dalam mengevaluasi performa model.

Hasil perbandingan kinerja algoritma menunjukkan bahwa *Isolation Forest* memiliki performa yang lebih baik dibandingkan *Local Outlier Factor* (LOF) dalam mendeteksi anomali pada dataset transaksi

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

digital. Isolation Forest mampu mendeteksi 37 dari 148 transaksi fraud dengan nilai recall sebesar 0,25 dan F1-score sebesar 0,2652, sedangkan LOF hanya mampu mendeteksi 2 transaksi fraud dengan nilai recall sebesar 0,0135 dan F1-score sebesar 0,0136. Hal ini menunjukkan bahwa pendekatan berbasis isolasi lebih efektif dibandingkan pendekatan berbasis kepadatan pada dataset dengan distribusi ekstrem.

Selain itu, hasil penelitian ini menegaskan bahwa pemilihan metrik evaluasi sangat penting dalam kasus dataset tidak seimbang. Metrik seperti *recall* dan F1-*score* lebih representatif dalam mengukur kemampuan model dalam mendeteksi anomali dibandingkan akurasi.

Secara keseluruhan, penelitian ini membuktikan bahwa *Isolation Forest* merupakan metode yang lebih stabil, efektif, dan relevan untuk diterapkan pada sistem deteksi fraud dalam transaksi digital yang memiliki karakteristik data besar, berdimensi tinggi, dan tidak seimbang. Dengan demikian, hipotesis penelitian yang menyatakan bahwa Isolation Forest memiliki performa lebih baik dibandingkan LOF dapat diterima.

4.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Optimalisasi Parameter Model

   Penelitian selanjutnya disarankan untuk melakukan tuning parameter secara lebih mendalam, seperti jumlah pohon (n_estimators) pada Isolation Forest dan jumlah tetangga (n_neighbors) pada LOF, guna memperoleh performa yang lebih optimal.
2. Penggunaan Teknik Penanganan Imbalanced Data

   Untuk meningkatkan kemampuan deteksi anomali, dapat diterapkan teknik penyeimbangan data seperti SMOTE, undersampling, atau hybrid sampling sehingga model dapat belajar pola anomali dengan lebih baik.
3. Pengembangan Model Hybrid

   Kombinasi antara metode unsupervised dan supervised learning (hybrid model) dapat dieksplorasi untuk meningkatkan akurasi dan kemampuan deteksi fraud, terutama dalam lingkungan data yang kompleks.
4. Penambahan Metrik Evaluasi

   Penelitian selanjutnya dapat menambahkan metrik evaluasi lain seperti AUC-ROC dan Precision-Recall Curve untuk memberikan gambaran performa model yang lebih komprehensif.
5. Uji pada Dataset dan Kasus Nyata

   Untuk meningkatkan validitas hasil, penelitian selanjutnya disarankan untuk menguji model pada dataset yang lebih beragam atau data transaksi real-time sehingga hasil penelitian lebih aplikatif dalam sistem industri.
6. Analisis Efisiensi Komputasi

   Perbandingan waktu komputasi dan efisiensi algoritma juga penting untuk dianalisis, terutama untuk implementasi pada sistem deteksi fraud berbasis real-time.

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026

5. UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi dan terima kasih kepada Tim Redaksi Jurnal Betrik (Besemah Teknologi Informasi dan Komputer) Institut Teknologi Pagar Alam atas dukungan dan kontribusinya dalam memfasilitasi penerbitan artikel ini.

DAFTAR RUJUKAN

[1] Y. Chen, C. Zhao, Y. Xu, C. Nie, and Y. Zhang, "Deep Learning in Financial Fraud Detection: Innovations, Challenges, and Applications," *Data Science and Management*, Aug. 2025, doi: 10.1016/j.dsm.2025.08.002.

[2] B. Fu, Y. Tong, Y. Li, Z. Tang, Z. Shang, and A. Li, "Financial Fraud Anomaly Detection of Listed Companies Based on Probabilistic Perspective Machine Learning Models," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 979–986. doi: 10.1016/j.procs.2025.08.121.

[3] S. O. Pinto and V. A. Sobreiro, "Literature review: Anomaly detection approaches on digital business financial systems," *Digital Business*, vol. 2, no. 2, Jan. 2022, doi: 10.1016/j.digbus.2022.100038.

[4] M. S. Mia, S. Roy, M. A. Ihsan, S. Hossain, and M. K. U. Ahamed, "Data-driven financial fraud detection using hybrid artificial and quantum intelligence," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 5, no. 4, Dec. 2025, doi: 10.1016/j.tbench.2025.100252.

[5] E. B. Satoto and Y. G. Wibowo, "Implementation of Artificial Intelligence in Fraud Detection and Prevention Through a Systematic Literature Review and Its Implications for the Financial Sector," *Ilomata International Journal of Management*, vol. 7, no. 1, pp. 460–483, Jan. 2026, doi: 10.61194/ijjm.v7i1.1919.

[6] Shanaa Mohammad and Abdallah sherief, "A Hybrid Anomaly Detection Framework Combining," Feb. 2026.

[7] E. Luzio and M. A. Ponti, "Real-Time Anomaly Detection with Synthetic Anomaly Monitoring (SAM)," Feb. 2025

[8] M. Sabuhi, M. Zhou, C.-P. Bezemer, and P. Musilek, "Applications of Generative Adversarial Networks in Anomaly Detection: A Systematic Literature Review," Oct. 2021.

[9] H. Xu, G. Pang, Y. Wang, and Y. Wang, "Deep Isolation Forest for Anomaly Detection," Jun. 2023, doi: 10.1109/TKDE.2023.3270293.

[10] S. A. Okolie, C. A. Amadi, J. N. Odii, E. C. Nwokorie, and U. .. C. Onyemauche, "Anomaly detection in heterogeneous cybersecurity data," *Franklin Open*, vol. 13, Dec. 2025, doi: 10.1016/j.fraope.2025.100426.

[11] E. F. Agyemang, "Anomaly detection using unsupervised machine learning algorithms: A simulation study," *Sci. Afr.*, vol. 26, Dec. 2024, doi: 10.1016/j.sciaf.2024.e02386.

[12] A. Goodge, B. Hooi, S. K. Ng, and W. S. Ng, "LUNAR: Unifying Local Outlier Detection Methods via Graph Neural Networks," Dec. 2021.

[13] T. Wu and Y. Wang, "Locally Interpretable One-Class Anomaly Detection for Credit Card Fraud Detection," Jun. 2022.

[14] M. De la Cruz Cabello, T. P. Sales, and M. R. Machado, "Log anomaly detection in AIOps: A real-world implementation using Large Language Models," *Systems and Soft Computing*, vol. 8, Jun. 2026, doi: 10.1016/j.sasc.2026.200475.

Received: 18-04-2026 | Accepted: 28-04-2026 | Published Online: 30-04-2026