# PENERAPAN DETEKSI ANOMALI MENGGUNAKAN MODEL *ISOLATION FOREST* UNTUK ANALISIS *FRAUD* PADA TRANSAKSI KEUANGAN AKUNTANSI

**Fadlillah Mukti Ayudewi [1], Ula Restu Rafifah [2]**
[1] Program Studi Teknologi Informasi, Universitas Aisyiyah Yogyakarta
[2] Program Studi Akuntansi, Universitas Aisyiyah Yogyakarta
Jl. Siliwangi (Ring Road Barat) No. 63, Nogotirto, Gamping, Sleman, Daerah Istimewa Yogyakarta
*fadlillahmuktiayudewi@unisayogya.ac.id*

**ABSTRAK**

Kecurangan dalam transaksi keuangan menjadi permasalahan yang dapat menimbulkan kerugian finansial serta menurunkan kepercayaan terhadap sistem keuangan. Seiring meningkatnya volume dan kompleksitas transaksi digital, metode deteksi fraud konvensional menjadi kurang efektif. Permasalahan utama dalam penelitian ini adalah bagaimana mendeteksi transaksi fraud pada dataset yang tidak seimbang serta mengidentifikasi anomali tanpa bergantung pada data berlabel yang terbatas. Penelitian ini bertujuan untuk menerapkan algoritma Isolation Forest sebagai metode anomaly detection dalam mendeteksi transaksi yang berpotensi sebagai fraud. Metode yang digunakan meliputi pengumpulan dan preprocessing dataset Credit Card Fraud Detection, pembagian data training dan testing, penerapan model Isolation Forest, serta evaluasi menggunakan confusion matrix, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model memiliki performa sangat baik dalam mengenali transaksi normal dengan precision dan recall mendekati 1,00. Namun, pada kelas fraud diperoleh precision sebesar 0,46, recall sebesar 0,29, dan F1-score sebesar 0,35, yang menunjukkan keterbatasan model akibat ketidakseimbangan dataset. Meskipun demikian, Isolation Forest memiliki potensi sebagai metode awal dalam mendeteksi anomali pada transaksi keuangan.

***Kata kunci :*** *fraud detection, anomaly detection, isolation forest, machine learning, transaksi keuangan.*

## 1. PENDAHULUAN

Kecurangan dalam transaksi keuangan *(fraud)* merupakan permasalahan yang krusial dalam dunia akuntansi dan sistem keuangan modern. Fraud tidak hanya menimbulkan kerugian finansial, tetapi juga dapat merusak reputasi perusahaan serta menurunkan kepercayaan pemangku kepentingan terhadap integritas laporan keuangan [1]. Seiring dengan meningkatnya digitalisasi sistem keuangan, dan transaksi elektronik, jumlah dan kompleksitas data transaksi juga semakin tinggi, sehingga proses identifikasi transaksi yang mencurigakan menjadi semakin sulit apabila hanya mengandalkan metode audit konvensional.

Dalam praktiknya, deteksi *fraud* masih banyak dilakukan menggunakan pendekatan manual atau berbasis aturan *(rule-based system).* Namun, metode tersebut memiliki keterbatasan karena hanya mampu mendeteksi pola *fraud* yang telah diketahui sebelumnya. Ketika muncul pola *fraud* baru yang tidak sesuai dengan aturan yang ada, sistem cenderung gagal dalam mengidentifikasinya [5]. Selain itu, tingginya volume data transaksi menyebabkan pendekatan manual menjadi tidak efisien dan memerlukan waktu yang lama [6] [7]. Namun, sebagian besar metode tersebut menggunakan pendekatan *supervised learning* yang membutuhkan data berlabel dalam jumlah besar. Pada kenyataannya, data *fraud* biasanya sangat terbatas dibandingkan data transaksi normal, sehingga terjadi ketidakseimbangan dataset *(class imbalance)* yang dapat menurunkan performa model [8].

Permasalahan utama dalam penelitian ini adalah bagaimana mendeteksi transaksi *fraud* pada dataset yang tidak seimbang serta bagaimana meningkatkan kemampuan model dalam mengidentifikasi transaksi yang menyimpang dari pola normal tanpa bergantung pada data berlabel dalam jumlah besar. Untuk mengatasi permasalahan tersebut, diperlukan pendekatan yang mampu mendeteksi anomali secara adaptif terhadap pola data yang kompleks.

Penelitian ini bertujuan untuk menerapkan algoritma *Isolation Forest* sebagai metode *anomaly detection* dalam mendeteksi transaksi yang berpotensi sebagai *fraud.* Algoritma ini dipilih karena mampu mengidentifikasi data yang menyimpang dari pola normal tanpa memerlukan data berlabel serta efektif digunakan pada dataset dengan tingkat ketidakseimbangan yang tinggi [9].

Rencana penyelesaian masalah dalam penelitian ini dilakukan melalui beberapa tahapan, yaitu studi literatur untuk memahami metode yang relevan, pengumpulan dan persiapan dataset transaksi keuangan, penerapan algoritma *Isolation Forest*, serta evaluasi model menggunakan metrik seperti *confusion matrix, precision, recall, dan F1-score.* Melalui tahapan tersebut, diharapkan model yang dihasilkan mampu mendeteksi anomali pada transaksi keuangan secara lebih efektif dan dapat digunakan sebagai alat bantu dalam analisis *fraud* pada sistem keuangan.

## 2. TINJAUAN PUSTAKA

### 2.1. *Fraud Detection*

*Fraud detection* merupakan proses identifikasi aktivitas yang mencurigakan atau tidak wajar dalam suatu sistem keuangan yang berpotensi menimbulkan kerugian bagi organisasi maupun individu. Dalam konteks sistem informasi akuntansi dan transaksi keuangan, *fraud detection* bertujuan untuk mengidentifikasi transaksi yang menyimpang dari pola transaksi normal sehingga dapat dilakukan tindakan pencegahan dan investigasi lebih lanjut [1].

*Fraud* dalam sistem keuangan dapat terjadi dalam berbagai bentuk, seperti manipulasi laporan keuangan, penyalahgunaan aset, transaksi tidak sah, maupun aktivitas pencucian uang. Seiring meningkatnya penggunaan teknologi digital dalam sistem keuangan, jumlah transaksi yang diproses oleh sistem juga meningkat secara signifikan. Kondisi ini menyebabkan proses pengawasan transaksi menjadi semakin kompleks karena auditor harus menganalisis data dalam jumlah besar dengan pola transaksi yang sangat beragam.

Pendekatan tradisional dalam *fraud detection* umumnya menggunakan metode audit manual atau sistem berbasis aturan *(rule-based system).* Pada pendekatan ini, suatu transaksi akan dikategorikan sebagai *fraud* apabila memenuhi kriteria tertentu yang telah ditentukan sebelumnya. Meskipun metode tersebut cukup efektif dalam mendeteksi pola *fraud* yang telah diketahui, sistem berbasis aturan memiliki keterbatasan dalam mendeteksi pola *fraud* baru yang belum pernah terjadi sebelumnya. Hal ini menyebabkan metode konvensional menjadi kurang efektif dalam menghadapi dinamika transaksi digital yang terus berkembang [5].

Perkembangan teknologi analisis data dan kecerdasan buatan memberikan peluang baru dalam meningkatkan efektivitas *fraud detection*. Dengan memanfaatkan teknik analisis data dan *machine learning,* sistem dapat mempelajari pola transaksi dari data historis serta mengidentifikasi aktivitas yang menyimpang dari pola tersebut. Pendekatan ini memungkinkan proses deteksi *fraud* dilakukan secara otomatis dan lebih adaptif terhadap perubahan pola transaksi.

### 2.2. *Machine Learning* dalam Deteksi *Fraud*

*Machine learning* merupakan salah satu pendekatan yang banyak digunakan dalam deteksi *fraud* karena kemampuannya dalam mempelajari pola data yang kompleks serta melakukan prediksi berdasarkan pola tersebut. Dalam konteks *fraud detection, machine learning* digunakan untuk mengklasifikasikan transaksi menjadi dua kategori utama yaitu transaksi normal dan transaksi *fraud*.

Berbagai algoritma machine learning telah diterapkan dalam penelitian deteksi *fraud,* seperti *Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, dan Neural Network* [12] [13]. Algoritma tersebut bekerja dengan memanfaatkan data historis yang telah diberi label untuk melatih model sehingga mampu mengenali pola yang membedakan transaksi normal dan transaksi *fraud*.

Pendekatan *supervised learning* sering digunakan karena memiliki tingkat akurasi yang cukup tinggi dalam proses klasifikasi. Misalnya, algoritma *Random Forest* mampu mengidentifikasi pola kompleks melalui kombinasi beberapa pohon keputusan, sementara *Neural Network* mampu menangkap hubungan *non-linear* dalam data transaksi yang sangat kompleks. Dengan memanfaatkan dataset pelatihan yang cukup besar, model *machine learning* dapat mengidentifikasi karakteristik transaksi yang mencurigakan secara lebih akurat dibandingkan metode berbasis aturan.

Namun demikian, penerapan *machine learning* dalam deteksi *fraud* juga menghadapi beberapa tantangan utama. Salah satu tantangan terbesar adalah ketidakseimbangan dataset *(class imbalance),* di mana jumlah transaksi *fraud* jauh lebih sedikit dibandingkan transaksi normal. Pada banyak dataset transaksi keuangan, proporsi transaksi fraud sering kali kurang dari satu persen dari seluruh data transaksi. Kondisi ini menyebabkan model *machine learning* cenderung lebih banyak mempelajari pola transaksi normal sehingga kemampuan dalam mendeteksi transaksi *fraud* menjadi terbatas [19].

Selain itu, beberapa algoritma *machine learning* memiliki kompleksitas model yang tinggi sehingga sulit untuk diinterpretasikan oleh auditor. Dalam praktik akuntansi forensik, interpretabilitas model menjadi aspek penting karena auditor perlu memahami alasan di balik suatu keputusan yang dihasilkan oleh sistem analisis data. Oleh karena itu, diperlukan metode yang tidak hanya memiliki tingkat akurasi yang baik tetapi juga mampu menangani dataset yang tidak seimbang serta dapat digunakan secara efektif dalam analisis transaksi keuangan.

### 2.3. *Anomaly Detection*

*Anomaly detection* merupakan teknik analisis data yang digunakan untuk mengidentifikasi data yang menyimpang dari pola normal dalam suatu dataset. Dalam konteks transaksi keuangan, anomali dapat diartikan sebagai transaksi yang memiliki karakteristik yang berbeda dibandingkan mayoritas transaksi lainnya.

Pendekatan *anomaly detection* banyak digunakan dalam berbagai bidang seperti keamanan jaringan, deteksi intrusi, monitoring sistem, serta deteksi *fraud* dalam sistem keuangan [18]. Metode ini menjadi sangat penting pada dataset yang memiliki ketidakseimbangan kelas yang tinggi, di mana data anomali atau *fraud* hanya merupakan sebagian kecil dari keseluruhan data transaksi.

Berbeda dengan metode klasifikasi tradisional yang memerlukan data berlabel, *anomaly detection* umumnya menggunakan pendekatan *unsupervised learning* yang tidak memerlukan label data dalam jumlah besar. Pendekatan ini memungkinkan sistem

untuk mempelajari distribusi data normal dan kemudian mengidentifikasi data yang memiliki penyimpangan signifikan dari distribusi tersebut [9][18].

Dalam konteks *fraud detection*, transaksi yang terdeteksi sebagai anomali dapat menjadi indikasi adanya aktivitas *fraud* yang perlu dianalisis lebih lanjut. Pendekatan ini sangat efektif karena *fraud* sering kali memiliki pola yang jarang terjadi dan berbeda dari transaksi normal. Dengan menggunakan metode *anomaly detection*, sistem dapat mengidentifikasi pola transaksi yang tidak biasa meskipun pola tersebut belum pernah ditemukan sebelumnya dalam dataset pelatihan.

Beberapa algoritma *anomaly detection* yang sering digunakan antara lain *One-Class Support Vector Machine (OC-SVM), Local Outlier Factor (LOF), Autoencoder,* dan *Isolation Forest.* Setiap algoritma memiliki karakteristik yang berbeda dalam mengidentifikasi anomali tergantung pada distribusi data dan kompleksitas dataset yang digunakan [9] [18].

### 2.4. *Isolation Forest*

*Isolation Forest* merupakan salah satu algoritma *anomaly detection* yang dirancang khusus untuk mendeteksi *outlier* dalam dataset besar secara efisien. Algoritma ini pertama kali diperkenalkan oleh Liu, Ting, dan Zhou sebagai metode yang memanfaatkan konsep isolasi data untuk mengidentifikasi anomali.

Berbeda dengan metode *anomaly detection* lainnya yang menggunakan pendekatan berbasis jarak atau kepadatan data, *Isolation Forest* bekerja dengan cara membangun sejumlah pohon isolasi secara acak. Pada setiap pohon, fitur dan nilai split dipilih secara acak untuk memisahkan data menjadi beberapa bagian. Data yang merupakan anomali cenderung lebih mudah terisolasi dibandingkan data normal karena memiliki karakteristik yang berbeda dari mayoritas data [23].

Proses isolasi tersebut menghasilkan *path length*, yaitu jumlah langkah yang diperlukan untuk mengisolasi suatu data dalam pohon. Data yang memiliki path length lebih pendek dianggap sebagai anomali karena lebih cepat terpisah dari data lainnya. Nilai *path length* tersebut kemudian digunakan untuk menghitung skor anomali yang menunjukkan tingkat keanehan suatu data dalam dataset [23].

*Isolation Forest* telah banyak digunakan dalam penelitian deteksi *fraud* pada transaksi keuangan. Beberapa penelitian menunjukkan bahwa metode ini mampu mendeteksi transaksi yang mencurigakan secara efektif pada dataset yang memiliki tingkat ketidakseimbangan yang tinggi. Penelitian sebelumnya menunjukkan bahwa *Isolation Forest* mampu meningkatkan kemampuan deteksi *fraud* pada transaksi kartu kredit dengan jumlah data yang sangat besar serta proporsi *fraud* yang sangat kecil [23]. Penelitian lain juga menunjukkan bahwa metode ini efektif digunakan untuk mendeteksi anomali pada data keuangan dan asuransi yang memiliki distribusi data yang tidak seimbang [25].

Beberapa keunggulan utama dari algoritma Isolation Forest antara lain:

a. Tidak memerlukan data berlabel, karena menggunakan pendekatan unsupervised learning sehingga dapat diterapkan pada dataset yang tidak memiliki label fraud.
b. Efisien untuk dataset berukuran besar, karena proses pembangunan pohon secara acak memiliki kompleksitas komputasi yang relatif rendah.
c. Mampu menangani data yang tidak seimbang, karena algoritma ini berfokus pada identifikasi anomali yang jarang terjadi dalam dataset.
d. Mampu mendeteksi pola anomali baru, karena tidak bergantung pada pola historis tertentu dalam data.

Dengan berbagai keunggulan tersebut, *Isolation Forest* menjadi salah satu metode yang banyak digunakan dalam penelitian deteksi *fraud* pada transaksi keuangan, khususnya pada dataset yang memiliki tingkat ketidakseimbangan yang tinggi [23] [25].

### 2.5. Penelitian terdahulu

Berbagai penelitian sebelumnya menunjukkan bahwa perkembangan teknologi *machine learning* telah memberikan kontribusi signifikan dalam meningkatkan efektivitas deteksi *fraud* pada transaksi keuangan. Sejumlah algoritma seperti *Logistic Regression, Decision Tree,* dan *Random Forest* telah banyak digunakan untuk mengklasifikasikan transaksi ke dalam kategori normal dan *fraud*, serta mampu memberikan tingkat akurasi yang cukup tinggi dalam berbagai studi [6] [7].

Selain itu, laporan dari *Association of Certified Fraud Examiners (ACFE)* menunjukkan bahwa kerugian akibat *fraud* di seluruh dunia mencapai miliaran dolar setiap tahunnya [2]. *Fraud* dapat terjadi dalam berbagai bentuk, seperti manipulasi laporan keuangan, penyalahgunaan aset, hingga transaksi keuangan yang tidak sah. Di Indonesia, hasil survei *ACFE* Indonesia *Chapter* menunjukkan bahwa kerugian akibat *fraud* mencapai sekitar Rp 242,26 miliar, dengan jenis *fraud* yang paling dominan adalah korupsi, diikuti oleh penyalahgunaan aset dan kecurangan laporan keuangan [3]. Studi kasus seperti skandal PT Tiga Pilar Sejahtera Food Tbk juga memperlihatkan bahwa *fraud* dapat memberikan dampak signifikan terhadap stabilitas keuangan organisasi [4].

Penelitian sebelumnya juga mengungkapkan bahwa metode deteksi *fraud* tradisional yang berbasis audit manual atau *rule-based system* memiliki keterbatasan dalam mendeteksi pola fraud yang baru dan kompleks. Metode ini umumnya hanya mampu mengidentifikasi pola yang telah diketahui sebelumnya, sehingga kurang efektif dalam

menghadapi dinamika transaksi digital yang terus berkembang [5]. Selain itu, meningkatnya volume transaksi menyebabkan pendekatan manual menjadi kurang efisien dan sulit diterapkan pada data skala besar.

Seiring dengan hal tersebut, berbagai penelitian menunjukkan bahwa pendekatan berbasis *machine learning* mampu mengatasi sebagian keterbatasan metode tradisional dengan memanfaatkan pola data historis untuk mendeteksi transaksi yang menyimpang. Namun demikian, sebagian besar metode yang digunakan masih berbasis *supervised learning* yang memerlukan dataset berlabel dalam jumlah besar. Pada praktiknya, data *fraud* cenderung sangat terbatas dibandingkan data transaksi normal, sehingga menyebabkan terjadinya ketidakseimbangan dataset *(class imbalance)* yang dapat menurunkan performa model dalam mendeteksi *fraud* [8].

Untuk mengatasi permasalahan tersebut, beberapa penelitian mulai mengadopsi pendekatan *anomaly detection* yang tidak memerlukan data berlabel dalam jumlah besar. Pendekatan ini bertujuan untuk mengidentifikasi data yang menyimpang dari pola normal. Salah satu algoritma yang banyak digunakan dalam penelitian adalah *Isolation Forest*, yang bekerja dengan cara mengisolasi data secara acak melalui pembentukan pohon isolasi [9]. Data yang merupakan anomali cenderung lebih cepat terisolasi dibandingkan data normal karena memiliki karakteristik yang berbeda dari mayoritas data.

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma *Isolation Forest* memiliki kemampuan yang baik dalam mendeteksi outlier pada dataset berukuran besar dan tidak seimbang [10] [11]. Hal ini menunjukkan bahwa metode tersebut memiliki potensi yang tinggi untuk diterapkan dalam deteksi fraud pada transaksi keuangan, khususnya pada kondisi data yang kompleks dan tidak seimbang.

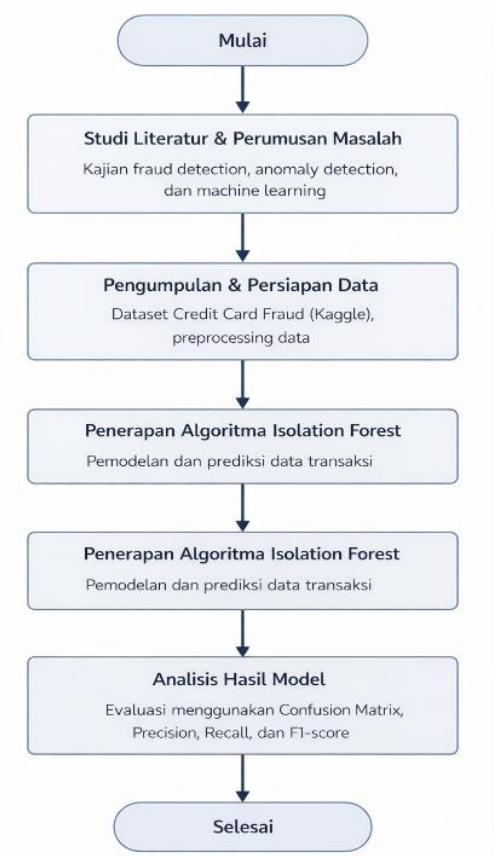
Berdasarkan kajian penelitian terdahulu penelitian ini berfokus pada penerapan algoritma *Isolation Forest* untuk mendeteksi anomali pada dataset transaksi keuangan serta mengevaluasi efektivitasnya dalam mengidentifikasi transaksi yang berpotensi sebagai *fraud.* Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode deteksi *fraud* berbasis *machine learning* serta menjadi alternatif pendekatan dalam analisis *fraud* pada bidang akuntansi forensik.

## 3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis analisis data dengan tujuan untuk menerapkan algoritma *Isolation Forest* dalam mendeteksi anomali pada transaksi keuangan. Pendekatan ini dipilih karena metode *anomaly detection* memungkinkan identifikasi pola transaksi yang menyimpang dari pola normal tanpa memerlukan data berlabel dalam jumlah besar. Hal ini menjadi penting dalam konteks *fraud detection*, di mana jumlah data *fraud* biasanya jauh lebih sedikit dibandingkan data transaksi normal.

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari dataset publik *Credit Card Fraud Detection* yang tersedia pada platform Kaggle. Dataset tersebut banyak digunakan sebagai *benchmark* dalam penelitian *fraud detection* karena memiliki jumlah data yang besar serta karakteristik ketidakseimbangan kelas yang tinggi antara transaksi normal dan transaksi fraud.

Proses penelitian dilakukan secara bertahap seperti yang ditunjukkan pada Gambar 1.




Gambar 1. Alur Penelitian Deteksi Fraud Menggunakan Algoritma Isolation Forest

### 3.1. Studi Literatur dan Perumusan Masalah

Tahap awal penelitian dilakukan dengan melakukan studi literatur terhadap berbagai penelitian yang berkaitan dengan *fraud detection, anomaly detection,* serta penerapan *machine learning* dalam analisis transaksi keuangan. Studi literatur ini bertujuan untuk memahami perkembangan metode yang telah digunakan dalam penelitian sebelumnya serta mengidentifikasi kelebihan dan keterbatasan dari masing-masing metode.

Dari hasil kajian literatur diketahui bahwa sebagian besar penelitian *fraud detection* menggunakan pendekatan *supervised learning* seperti *Random Forest*, *Logistic Regression*, dan *Gradient Boosting.* Meskipun metode tersebut memiliki tingkat akurasi yang tinggi, pendekatan tersebut memerlukan

dataset berlabel yang cukup besar. Pada kenyataannya, data *fraud* sering kali sangat terbatas sehingga menyebabkan ketidakseimbangan dataset.

Berdasarkan permasalahan tersebut, penelitian ini menggunakan pendekatan *unsupervised anomaly detection* dengan algoritma *Isolation Forest* untuk mendeteksi transaksi yang menyimpang dari pola normal.

### 3.2. Pengumpulan dan Persiapan Data

Dataset yang digunakan dalam penelitian ini adalah *Credit Card Fraud Detection Dataset* yang tersedia secara publik pada platform Kaggle Dataset ini berisi data transaksi kartu kredit yang telah diproses menggunakan teknik *Principal Component Analysis (PCA)* untuk menjaga kerahasiaan informasi pengguna.

Dataset tersebut terdiri dari lebih dari 280.000 transaksi, dengan proporsi transaksi *fraud* sekitar 0,17% dari seluruh data. Kondisi ini menunjukkan bahwa dataset memiliki tingkat ketidakseimbangan yang sangat tinggi antara transaksi normal dan transaksi *fraud.*

Tahapan persiapan data meliputi beberapa langkah berikut:

a. *Import* dataset menggunakan library Python seperti Pandas.
b. Eksplorasi dataset untuk memahami struktur data, jumlah fitur, serta distribusi kelas.
c. Penghapusan atribut yang tidak relevan, seperti kolom *Time dan Amount.*
d. Penentuan fitur (X) dan target (y), di mana fitur berisi seluruh variabel kecuali label class.
e. Pembagian dataset menjadi data *training* dan *testing* dengan rasio 70:30 menggunakan metode *train-test split.*

Tahapan *preprocessing* ini bertujuan untuk memastikan bahwa data yang digunakan dalam proses pelatihan model telah bersih dan siap digunakan untuk analisis.

### 3.3. Penerapan Algoritma

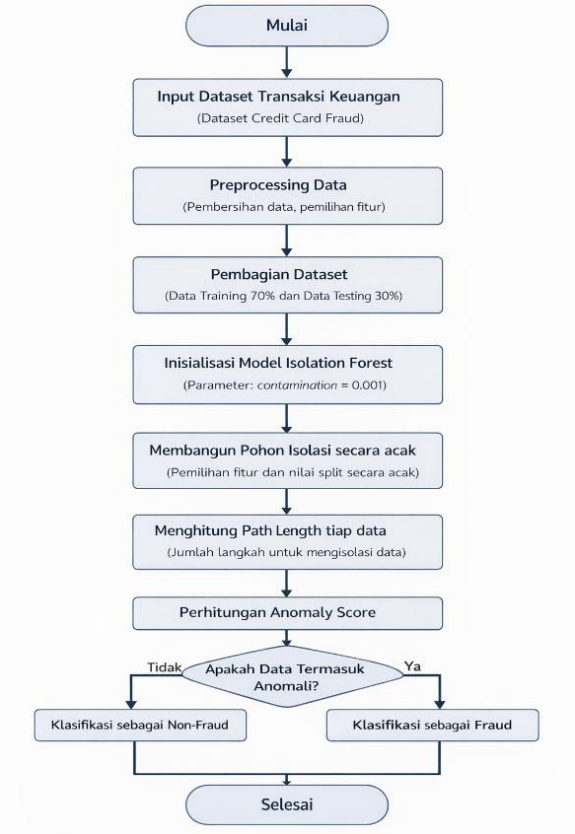
Model yang digunakan dalam penelitian ini adalah yang diimplementasikan menggunakan *library Scikit-learn* pada bahasa pemrograman Python.

*Isolation Forest* merupakan algoritma berbasis ensemble yang bekerja dengan cara membangun sejumlah pohon isolasi secara acak. Setiap pohon dibangun dengan memilih fitur dan nilai split secara acak sehingga proses isolasi data dapat dilakukan secara efisien. Data yang merupakan anomali akan membutuhkan jumlah pemisahan yang lebih sedikit dibandingkan data normal dengan menggunakan *parameter contamination* = 0.001 dan *random_state* = 42.

Model kemudian dilatih menggunakan data *training* untuk mempelajari pola transaksi normal. Setelah proses pelatihan selesai, model digunakan untuk melakukan prediksi terhadap data testing.

Hasil prediksi dari *Isolation Forest* berupa nilai -1 untuk anomali dan 1 untuk data normal. Nilai tersebut kemudian dikonversi menjadi label klasifikasi biner yaitu *fraud* dan *non-fraud* untuk memudahkan proses evaluasi model.

Alur model deteksi *fraud* menggunakan algoritma *isolation forest* disajikan pada gambar 2




Gambar 2. Tahapan Model Deteksi *Fraud* Menggunakan Algoritma *Isolation Forest*

### 3.4. Analisis

Tahap analisis dilakukan untuk mengevaluasi performa model dalam mendeteksi transaksi *fraud.* Evaluasi dilakukan menggunakan beberapa metrik evaluasi klasifikasi yang umum digunakan dalam penelitian *machine learning*, yaitu: *Confusion Matrix, Precision, Recall, F1-score.*

*Confusion Matrix* digunakan untuk mengetahui jumlah prediksi yang benar dan salah yang dihasilkan oleh model. Sementara itu, *precision* digunakan untuk mengukur tingkat ketepatan model dalam memprediksi transaksi *fraud*, sedangkan *recall* digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh transaksi *fraud* yang ada.

*F1-score* digunakan sebagai metrik gabungan antara precision dan recall untuk memberikan gambaran yang lebih seimbang mengenai performa model.

Hasil evaluasi tersebut kemudian dianalisis untuk mengetahui efektivitas algoritma *Isolation Forest* dalam mendeteksi anomali pada transaksi keuangan

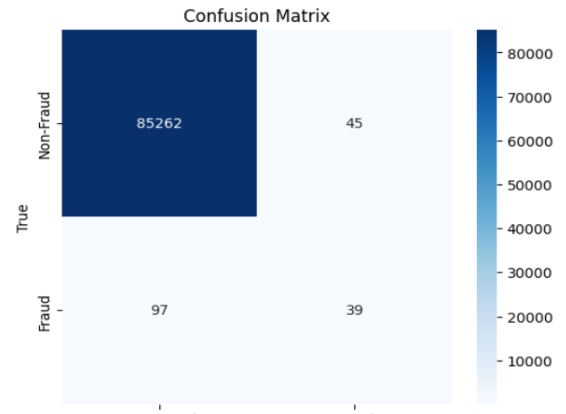
serta untuk mengidentifikasi faktor-faktor yang mempengaruhi kinerja model.

## 4. HASIL DAN PEMBAHASAN

Penelitian ini mengimplementasikan metode deteksi anomali menggunakan algoritma *Isolation Forest* dengan memanfaatkan bahasa pemrograman Python dan pustaka *machine learning scikit-learn*. Dataset yang digunakan adalah *Credit Card Fraud Detection* yang berisi data transaksi kartu kredit yang telah melalui proses transformasi menggunakan teknik *Principal Component Analysis (PCA)* untuk menjaga kerahasiaan identitas pengguna. Dataset ini memiliki karakteristik ketidakseimbangan kelas yang tinggi, di mana jumlah transaksi *fraud* hanya sekitar 0,17% dari seluruh data transaksi.

Tahap analisis dimulai dengan proses *preprocessing* data yang meliputi eksplorasi dataset, pemilihan atribut yang relevan, serta pembagian dataset menjadi data pelatihan *(training data)* dan data pengujian *(testing data)* dengan rasio 70:30. Setelah itu, algoritma *Isolation Forest* diterapkan untuk mengidentifikasi transaksi yang menyimpang dari pola transaksi normal. Metode ini bekerja dengan membangun pohon isolasi secara acak sehingga data yang memiliki karakteristik berbeda dari mayoritas data akan lebih cepat terisolasi dan dikategorikan sebagai anomali.

Hasil pemodelan kemudian dievaluasi menggunakan beberapa metrik evaluasi klasifikasi yaitu *confusion matrix, precision, recall*, dan *F1-score* untuk mengetahui kemampuan model dalam mendeteksi transaksi *fraud*. Berdasarkan hasil pengujian model *Isolation Forest*, diperoleh confusion matrix yang ditunjukkan pada Gambar 3.




Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 1.00 | 1.00 | 85307 |
| 1 | 0.46 | 0.29 | 0.35 | 136 |
| accuracy | | | 1.00 | 85443 |
| macro avg | 0.73 | 0.64 | 0.68 | 85443 |
| weighted avg | 1.00 | 1.00 | 1.00 | 85443 |

Gambar 3. *Confusion Matrix* Hasil Prediksi Model *Isolation Forest*

Berdasarkan hasil pengujian model *Isolation Forest*, diperoleh *confusion matrix* yang menggambarkan kinerja model dalam mengklasifikasikan transaksi ke dalam dua kategori yaitu *non-fraud* dan *fraud*. *Confusion matrix* digunakan untuk menunjukkan jumlah prediksi yang benar maupun salah yang dihasilkan oleh model pada setiap kelas.

Hasil *confusion matrix* menunjukkan bahwa model berhasil mengklasifikasikan 85.262 transaksi *non-fraud* dengan benar yang termasuk dalam kategori *True Negative (TN).* Selain itu, terdapat 45 transaksi *non-fraud* yang salah diklasifikasikan sebagai *fraud*, yang termasuk dalam kategori *False Positive (FP).*

Pada kelas fraud, model mampu mendeteksi 39 transaksi *fraud* secara benar yang termasuk dalam kategori *True Positive (TP).* Namun demikian, masih terdapat 97 transaksi *fraud* yang tidak berhasil dideteksi oleh model, yang termasuk dalam kategori *False Negative (FN).* Hasil ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali transaksi normal, namun masih memiliki keterbatasan dalam mendeteksi seluruh transaksi *fraud* yang terdapat dalam dataset.

Selain menggunakan *confusion matrix*, evaluasi performa model juga dilakukan menggunakan *classification report* yang mencakup metrik *precision, recall*, dan *F1-score.* Metrik ini digunakan untuk memberikan gambaran yang lebih komprehensif mengenai kemampuan model dalam mendeteksi transaksi *fraud*.

Berdasarkan hasil evaluasi model, diperoleh nilai *precision* sebesar 1.00 untuk kelas *non-fraud*, yang menunjukkan bahwa hampir seluruh transaksi yang diprediksi sebagai *non-fraud* benar-benar merupakan transaksi normal. *Nilai recall* untuk kelas *non-fraud* juga mencapai 1.00, yang berarti model mampu mengenali hampir seluruh transaksi normal dalam dataset.

Namun pada kelas *fraud*, nilai *precision* sebesar 0.46 menunjukkan bahwa dari seluruh transaksi yang diprediksi sebagai *fraud*, sekitar 46% benar-benar merupakan transaksi *fraud*. Sementara itu, nilai *recall* sebesar 0.29 menunjukkan bahwa model hanya mampu mendeteksi sekitar 29% dari seluruh transaksi *fraud* yang terdapat dalam dataset. Nilai *F1-score* sebesar 0.35 pada kelas *fraud* menunjukkan bahwa performa model dalam mendeteksi transaksi *fraud* masih relatif rendah dibandingkan dengan kelas *non-fraud.*

Performa model yang rendah pada kelas *fraud* terutama disebabkan oleh karakteristik dataset yang memiliki ketidakseimbangan kelas *(class imbalance)* yang sangat tinggi, di mana jumlah transaksi normal jauh lebih banyak dibandingkan transaksi *fraud*. Kondisi ini menyebabkan model lebih banyak mempelajari pola transaksi normal sehingga kemampuan dalam mendeteksi transaksi *fraud* menjadi terbatas. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa

ketidakseimbangan dataset merupakan salah satu faktor utama yang mempengaruhi performa model dalam deteksi *fraud* berbasis *machine learning* [12] [19].

Meskipun demikian, penggunaan algoritma *Isolation Forest* tetap memiliki keunggulan karena mampu mendeteksi anomali tanpa memerlukan data berlabel dalam jumlah besar. Metode ini sangat relevan untuk diterapkan pada sistem deteksi *fraud* yang memiliki jumlah transaksi sangat besar serta proporsi *fraud* yang relatif kecil.

Dalam konteks akuntansi forensik dan audit berbasis data, algoritma *Isolation Forest* dapat digunakan sebagai alat bantu analisis awal untuk mengidentifikasi transaksi yang memiliki pola tidak biasa. Transaksi yang terdeteksi sebagai anomali kemudian dapat dianalisis lebih lanjut oleh auditor atau analis keuangan untuk memastikan apakah transaksi tersebut benar-benar merupakan aktivitas fraud atau hanya merupakan variasi transaksi yang tidak umum.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa penerapan metode *anomaly detection* berbasis *machine learning* memiliki potensi yang cukup besar dalam meningkatkan efektivitas proses deteksi *fraud* pada sistem transaksi keuangan modern.

## 5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *Isolation Forest* dapat digunakan untuk mendeteksi anomali pada transaksi keuangan sebagai indikasi potensi *fraud*. Hasil evaluasi model menunjukkan bahwa model memiliki performa yang sangat baik dalam mengenali transaksi normal dengan nilai *precision* dan *recall* yang tinggi. Namun, kemampuan model dalam mendeteksi transaksi *fraud* masih terbatas yang ditunjukkan oleh nilai *precision* sebesar 0.46, *recall* sebesar 0.29, dan *F1-score* sebesar 0.35. Hal ini disebabkan oleh karakteristik dataset yang memiliki tingkat ketidakseimbangan kelas yang tinggi, di mana jumlah transaksi normal jauh lebih banyak dibandingkan transaksi *fraud*.

Meskipun demikian, metode *anomaly detection* menggunakan *Isolation Forest* tetap memiliki potensi untuk digunakan sebagai alat bantu dalam mendeteksi transaksi yang mencurigakan pada sistem transaksi keuangan. Untuk penelitian selanjutnya disarankan melakukan optimasi parameter model, menerapkan teknik penyeimbangan data seperti *oversampling* atau *undersampling*, serta mengkombinasikan metode *anomaly detection* dengan algoritma *machine learning* lainnya agar kemampuan model dalam mendeteksi transaksi *fraud* dapat meningkat.

## DAFTAR PUSTAKA

[1] Y. Zhang, S. Zhang, J. Wang, and L. Wang, “Financial fraud detection: A literature review,” Humanities and Social Sciences Communications, vol. 11, no. 1, pp. 1–16,2024.

[2] Association of Certified Fraud Examiners (ACFE), “Report to the Nations: Global Study on Occupational Fraud and Abuse,” 2022.

[3] M. T. Arrahman and N. M. Machdar, “Pengaruh Fraud Pentagon dan Asimetri Informasi terhadap Fraudulent Financial Statement dengan Kualitas Laba sebagai Pemoderasi,” JURA: Jurnal Riset Akuntansi, vol. 2, no. 1, pp. 82–94, Feb. 2024, doi: 10.54066/jura-itb.v2i1.1322.

[4] Universitas Bina Nusantara, "https://accounting.binus.ac.id/," Bina Nusantara, [Online]. Available: https://accounting.binus.ac.id/2021/12/27/kasus-fraud-pt-tiga-pilar-sejahtera-masalah-fraud/. [Accessed 04 12 2025].

[5] A. Srivastava and A. Sahrawat, “Challenges in fraud detection using traditional audit,” International Journal of Accounting Research, vol. 12, no. 3, pp. 45–53, 2021.

[6] G. Wang, “Customer segmentation in the digital marketing using a Q-learning based differential evolution algorithm integrated with K-means clustering,” PLOS ONE, vol. 20, no. 3, pp. 1–17, 2025.

[7] R. Chalapathy and S. Chawla, “Deep learning for anomaly detection: A survey,” ACM Computing Surveys, vol. 54, no. 2, pp. 1–38, 2021.

[8] J. Lin, “Application of machine learning in predicting consumer behavior for targeted marketing strategies,” PLOS ONE, vol. 20, no. 2, pp. 1–13, 2025.

[9] K. Sharma and M. Kumar, “Unsupervised learning methods for anomaly detection in financial datasets: A case study,” Inf. Syst. Front., vol. 25, pp. 1345–1361, 2023.

[10] M. Z. A. Bakar, M. F. Zolkipli, and A. S. Malik, “Performance analysis of Isolation Forest in fraud detection,” ResearchGate Preprint, 2025.

[11] X. Xu, C. Sun, Y. Li, and T. Wang, “Hybrid anomaly detection framework for imbalanced financial data,” F1000Research, vol. 14, no. 233, pp. 1–12, 2025.

[12] M. Sopiyan, Fauziah, and Y. F. Wijaya, “Fraud Detection Using Random Forest Classifier, Logistic Regression, and Gradient Boosting Classifier Algorithms on Credit Cards,” JUITA: Jurnal Informatika, vol. 10, no. 1, May 2022.

[13] R. Permana et al., “Analisis Penggunaan Machine Learning untuk Deteksi Penipuan di Sektor Keuangan: Tinjauan Literatur,” TEKNOBIS: Teknologi, Bisnis dan Pendidikan, vol. 2, no. 2, pp. 244–248, Sep. 2024.

[14] G. A. Ohyver, J. T. Beng, D. Trisnawarman, S. Tiatri, and D. A. Ohyver, “Analisis Data Keuangan Menggunakan Isolation Forest untuk Meningkatkan Akurasi Deteksi Anomali,” Journal of Information Technology and Computer Science (INTECOMS), vol. 8, no. 6, 2025.

[15] C. D. R. Amirillah, "Deteksi Transaksi Penipuan pada Sektor Perbankan Menggunakan Ruled-Based Model dan Pembelajaran Mesin," Jurnal Nasional Teknik Elektro dan Teknologi Informasi, vol. 14, no. 2, May 2025.
[16] L. Patel and D. Shah, "Machine learning-based anomaly detection in online payments," J. Inf. Secur. Appl., vol. 70, pp. 103–146, 2022.
[17] H. Pourhabibi, Z. Ong, B. K. Boo, and Y. Y. Choong, "Fraud detection: A systematic literature review of graph-based anomaly detection approaches," Decision Support Systems, vol. 133, pp. 113–303, 2020.
[18] Y. Li, C. Sun, and X. Xu, "Anomaly detection in financial transactions using unsupervised learning," Information Sciences, vol. 592, pp. 200–212, 2022.
[19] H. Song and J. Huang, "Evaluating anomaly detection models for highly imbalanced financial datasets," Expert Syst. Appl., vol. 201, pp. 117–153, 2022.
[20] M. West and J. Bhattacharya, "Outlier detection in imbalanced datasets: Comparative study of anomaly detection algorithms," Appl. Soft Comput., vol. 115, pp. 108–210, 2022.
[21] S. Verma and A. Shukla, "Comparative analysis of anomaly detection techniques for financial fraud," Procedia Comput. Sci., vol. 192, pp. 4120–4129, 2021.
[22] Y. Wang, H. Zhou, and L. Chen, "Application of unsupervised learning for fraud detection in credit card transactions," IEEE Access, vol. 10, pp. 115423–115433, 2022.
[23] D. Jin and Q. Wu, "Financial fraud detection using hybrid feature selection and Isolation Forest," Future Gener. Comput. Syst., vol. 137, pp. 234–246, 2022.
[24] S. Karim, R. Qamar, and F. Azam, "A hybrid model for credit card fraud detection using clustering and classification," Complex Intell. Syst., vol. 9, pp. 21–34, 2023.
[25] T. Nguyen and S. Duong, "Application of Isolation Forest in anomaly detection for insurance fraud," J. Big Data, vol. 9, no. 1, pp. 1–15, 2022.