Terbit online pada laman web jurnal: **https://ejurnalunsam.id/index.php/jicom/**

JICOM
Jurnal Informatika dan Teknologi Komputer
Vol. 07 No. 01 (2026) 14 - 22 E-ISSN :2774-7115 P-ISSN: 2775-2089

Identifikasi Pola Fraud pada Ekosistem Pembayaran Digital menggunakan Metode Isolation Forest

M. Willi Akbar[1], Septiani Kusuma Ningrum[2], Mira Afrina[3], Ali Ibrahim[4*]
Fakultas Ilmu Komputer, Universitas Sriwijaya
[1]M.willy.akbar@gmail.com, [2]septianikuning@gmail.com, [3]miraafrina@unsri.ac.id, [4]aliibrahim@unsri.ac.id*

Abstract

The transition to a digital economy in Indonesia faces a crucial challenge in the form of increasingly sophisticated fraud attacks. This study proposes an unsupervised learning approach to recognize patterns of new fraud attacks as a basis for strengthening institutional supervision capacity. This study focuses on identifying anomalies without relying on existing historical labels by utilizing the Isolation Forest ensemble algorithm. The model successfully maps suspicious transaction characteristics thanks to the application of deep feature engineering, which includes behavioral analysis, address correlation, and signal extraction from IP. The evaluation results show that this unsupervised approach is capable of identifying 18% of the total labeled fraud cases, proving its relevance in capturing actual attack signals. More importantly, a qualitative analysis of the anomalies found successfully characterized a Silent Attack Pattern, which is a combination of multi-factor risks that could potentially be overlooked by conventional detection systems. This finding gives opportunities to regulatory institutions to shift from reactive supervision to proactive threat discovery, which ultimately supports the creation of a secure and sustainable digital financial ecosystem in line with the Sustainable Development Goals (SDGs).

Keywords: Fraud Detection, Ensemble Learning, Isolation Forest, Unsupervised Learning, Institutional Supervision, Digital Economy

Abstrak

Transformasi menuju ekonomi digital di Indonesia dihadapkan pada tantangan krusial berupa meningkatnya serangan fraud yang semakin canggih. Penelitian ini mengajukan sebuah pendekatan unsupervised learning untuk mengenali pola serangan fraud generasi baru sebagai dasar penguatan kapasitas supervisi institusional. Penelitian ini berfokus pada identifikasi anomali tanpa bergantung pada label historis yang ada dengan memanfaatkan algoritma ensemble Isolation Forest. Model berhasil memetakan karakteristik transaksi yang mencurigakan berkat penerapan rekayasa fitur yang mendalam, yang mencakup analisis perilaku, korelasi alamat, dan ekstraksi sinyal dari IP. Hasil evaluasi menunjukkan bahwa pendekatan unsupervised ini mampu mengidentifikasi 18% dari total kasus fraud yang telah dilabelkan, membuktikan relevansinya dalam menangkap sinyal serangan yang sesungguhnya. Lebih penting lagi, analisis kualitatif terhadap anomali yang ditemukan berhasil mengkarakterisasi sebuah Pola Serangan Senyap, yaitu kombinasi multi-faktor risiko yang berpotensi terlewatkan oleh sistem deteksi konvensional. Temuan ini menyajikan sebuah wawasan baru bagi institusi regulator untuk beralih dari supervisi reaktif ke penemuan ancaman proaktif, yang pada akhirnya mendukung terciptanya ekosistem keuangan digital yang aman dan berkelanjutan sejalan dengan Tujuan Pembangunan Berkelanjutan (SDGs).

Kata kunci: Fraud Detection, Ensemble Learning, Isolation Forest, Unsupervised Learning, Institutional Supervision, Digital Economy

1 Pendahuluan

Perkembangan pesat teknologi digital dalam satu dekade terakhir telah mentransformasi lanskap sektor keuangan Indonesia secara fundamental. Laporan Bank Indonesia menunjukkan volume transaksi pembayaran digital melonjak 287% dalam lima tahun, mencapai Rp 52.000 triliun pada tahun 2022 [1]. Pertumbuhan eksponensial ini dimungkinkan oleh sebuah arsitektur sistem pembayaran digital nasional yang kompleks, yang terdiri dari infrastruktur inti BI-FAST, layer interkoneksi Gerbang Pembayaran Nasional (GPN), dan layer aplikasi berbasis Open Banking [2]. Ekosistem dinamis ini memiliki dua pilar regulasi utama yaitu Bank Indonesia sebagai operator sistemik dan regulator keamanan, serta Otoritas Jasa Keuangan (OJK) yang fokus pada pengawasan penyedia jasa keuangan digital

Diterima Redaksi : 25-02-2026 | Selesai Revisi : 14-03-2026 | Diterbitkan Online : 01-04-2026

dan perlindungan konsumen. Di balik kemudahan dan efisiensi transaksi real-time ini, muncul risiko yang sepadan. Data OJK mencatat kerugian akibat fraud digital mencapai Rp 1,2 triliun hanya dalam semester pertama 2023, dengan tingkat pertumbuhan kasus tahunan sebesar 45% [3].

Ancaman fraud digital ini bersifat multidimensi dan dampaknya mengancam keberlanjutan pertumbuhan ekonomi. Pada level makro, fenomena ini tidak hanya menimbulkan kerugian finansial langsung, tetapi juga secara signifikan menggerus kepercayaan publik yang merupakan fondasi dari ekonomi digital. Penurunan kepercayaan masyarakat tercatat mencapai 30% setelah maraknya kasus penipuan [4] dan secara langsung meningkatkan biaya modal di sektor fintech sebesar 120 basis poin akibat meningkatnya premi risiko [5]. Bagi institusi regulator, maraknya fraud merupakan tantangan langsung terhadap mandat mereka dalam menjaga stabilitas sistem keuangan. Kompleksitas tantangan dalam menanggulangi ancaman ini terletak pada tiga dimensi utama. Pertama, evolusi pola serangan yang semakin canggih, seperti deepfake social engineering dan multi-vector fraud yang memanfaatkan kerentanan lintas platform [6]. Kedua, karakteristik data transaksi yang sangat tidak seimbang (imbalanced), di mana aktivitas fraud merupakan minoritas yang sangat kecil namun berpotensi menimbulkan kerugian massif. Ketiga, tuntutan deteksi real-time dengan tingkat false positive yang minimal untuk tidak mengganggu kenyamanan jutaan transaksi yang sah [7].

Tantangan deteksi fraud yang kompleks telah mendorong pengembangan berbagai pendekatan solusi. Secara historis, industri keuangan mengandalkan sistem berbasis aturan (rule-based systems) yang menggunakan logika statis, seperti memblokir transaksi di atas nominal tertentu. Pendekatan ini kemudian dilengkapi dengan metode statistik untuk analisis anomali. Kedua metode konvensional ini, bagaimanapun, memiliki kelemahan fundamental yaitu mencakup sifatnya yang kaku, sulit beradaptasi dengan modus operandi serangan yang cepat berubah, dan cenderung menghasilkan tingkat false positive yang tinggi. Keterbatasan inilah yang mendorong pergeseran paradigma ke arah solusi yang lebih dinamis dan adaptif, yaitu Machine Learning (ML).

Penelitian canggih di bidang ML telah membuktikan keunggulannya dalam mendeteksi pola yang kompleks. Berbagai arsitektur telah diusulkan, mulai dari penggunaan Graph Neural Networks (GNN) untuk membongkar pola fraud kolusif [8], aplikasi Federated Learning untuk memungkinkan pelatihan model antar institusi tanpa mengorbankan privasi data [7], hingga teknik penanganan data tidak seimbang seperti SMOTE [9]. Di antara beragam pendekatan ini, metode ensemble learning secara konsisten muncul sebagai solusi yang sangat menjanjikan. Pendekatan ensemble terbukti mampu meningkatkan akurasi, menurunkan tingkat kesalahan, dan menjaga latensi tetap rendah melalui kombinasi kekuatan dari berbagai algoritma, yang menjadikannya kandidat ideal untuk sistem deteksi fraud kontemporer [10].

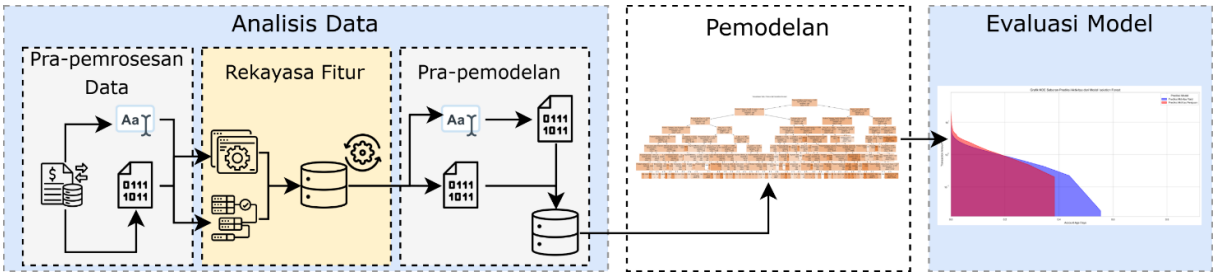
Sebagian besar metode ensemble yang ada bersifat supervised dan bergantung pada data berlabel, sehingga hanya mampu mengenali pola fraud yang telah diketahui sebelumnya. Guna mengatasi keterbatasan ini dan menemukan ancaman generasi baru, penelitian ini mengajukan pendekatan unsupervised learning. Secara spesifik, algoritma Isolation Forest dipilih karena paradigmanya yang unik. Berbeda dengan model lain, Isolation Forest tidak mencari pola fraud secara spesifik, melainkan bekerja dengan cara mengisolasi observasi data yang secara matematis paling berbeda [11]. Hipotesis utama penelitian ini adalah bahwa metode ini tidak hanya akan mengidentifikasi fraud yang sudah dilabelkan, tetapi juga berpotensi menemukan pola-pola serangan baru yang terlewatkan oleh sistem pelabelan konvensional. Inilah inti dari Pengenalan Pola yang menjadi tujuan utama penelitian ini.

Penelitian ini pada akhirnya tidak hanya memberikan kontribusi teknis, tetapi juga memiliki implikasi strategis yang lebih luas. Kemampuan untuk mengenali pola fraud baru secara proaktif merupakan dasar esensial bagi penguatan kapasitas supervisi institusional, yang memungkinkan regulator untuk bertindak berdasarkan intelijen data, bukan hanya laporan reaktif. Lebih jauh, penelitian ini bersifat multidisiplin dengan menunjukkan bagaimana ilmu komputer dan keamanan siber dapat memberikan solusi konkret untuk tantangan di bidang ekonomi. Keberhasilan dalam menciptakan ekosistem keuangan digital yang lebih aman secara langsung akan menyumbang pada pencapaian Tujuan Pembangunan Berkelanjutan (Sustainable Development Goals), khususnya pada SDG 8 (Pekerjaan Layak dan Pertumbuhan Ekonomi) dengan mendorong kepercayaan dan partisipasi dalam ekonomi digital, serta SDG 9 (Industri, Inovasi, dan Infrastruktur) dengan membangun infrastruktur keuangan yang tangguh dan inovatif [12]. Penelitian ini dengan demikian diposisikan sebagai sebuah langkah strategis untuk mendukung transformasi ekonomi nasional yang aman dan berkelanjutan.

2 Metode Penelitian

Metodologi penelitian merupakan kerangka kerja sistematis yang menjadi landasan bagi pelaksanaan penelitian ini. Bab ini menguraikan secara rinci pendekatan, sumber data, serta serangkaian tahapan teknis. Alur metodologi ini dirancang secara cermat, dimulai dari persiapan data mentah, rekayasa fitur untuk mengekstrak sinyal-sinyal tersembunyi, hingga penerapan model ensemble learning tanpa pengawasan

(unsupervised) untuk menemukan aktivitas mencurigakan. Setiap tahapan dilaksanakan dengan tujuan untuk membangun sebuah analisis yang tidak hanya prediktif, tetapi juga mampu memberikan wawasan baru yang berpotensi melampaui kapabilitas sistem pelabelan yang ada saat ini. Metodologi penelitian ini ditampilkan dalam gambar 1 dibawah ini.



Gambar 1. Metodologi Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif. Pendekatan ini dipilih karena fokus utama penelitian adalah analisis data numerik dalam volume besar dan penerapan pemodelan komputasi untuk mengidentifikasi, mengukur, dan memvalidasi pola-pola yang signifikan secara statistik. Ruang lingkup penelitian ini bersifat non-geografis, dengan objek studi yang berpusat pada analisis dan pemodelan data transaksional yang merepresentasikan aktivitas dalam ekosistem ekonomi digital.

2.1 Sumber dan pengumpulan data

Penelitian ini memanfaatkan data sekunder yang tersedia untuk publik dan dapat diakses melalui platform repositori data Kaggle [13]. Dataset yang digunakan adalah simulasi data transaksi dari platform e-commerce, yang secara spesifik dirancang untuk studi deteksi penipuan. Dataset ini terdiri dari 1.472.952 baris data transaksi dengan 16 fitur (kolom) awal. Salah satu fitur kunci dalam dataset ini adalah Is Fraudulent, sebuah label biner (0 untuk sah, 1 untuk fraud) yang telah ditentukan oleh pengembang dataset. Namun, sesuai dengan tujuan penelitian untuk menemukan potensi serangan baru, label Is Fraudulent ini tidak digunakan dalam proses pelatihan model. Label ini secara strategis hanya akan digunakan pada tahap akhir sebagai pembanding eksternal untuk memvalidasi relevansi pola anomali yang ditemukan oleh model. Pendekatan ini memungkinkan penelitian untuk menguji hipotesis bahwa ada potensi aktivitas mencurigakan lain yang tidak teridentifikasi oleh sistem pelabelan asli.

2.2 Tahapan analisis data

Alur analisis data dalam penelitian ini dibagi menjadi beberapa tahapan utama yang sistematis, dari persiapan awal hingga siap untuk pemodelan.

2.2.1 Prapemrosesan Data

Tahap awal ini berfokus pada eksplorasi dan pembersihan data mentah untuk memastikan kualitas dan integritas data sebelum dilakukan analisis lebih lanjut. Mengingat data mentah berpotensi mengandung noise atau inkonsistensi, prapemrosesan menjadi fondasi penting. Cakupan utama pada tahap ini adalah melakukan analisis deskriptif untuk memahami karakteristik dasar setiap fitur, termasuk identifikasi tipe data (numerik dan kategorikal) serta distribusi nilainya. Tahap ini memastikan bahwa dataset yang akan digunakan dalam rekayasa fitur dan pemodelan berada dalam kondisi optimal [14].

2.2.2 Rekayasa Fitur

Tahap ini merupakan inti dari proses pengayaan data, di mana penelitian ini membangun fitur-fitur baru yang informatif dari fitur asli untuk memperkuat kemampuan model dalam mengenali pola [15]. Proses ini didasarkan pada hipotesis mengenai perilaku transaksional yang mencurigakan. Proses rekayasa fitur mencakup tahapan dibawah ini yaitu,

1. Dekomposisi Fitur Waktu

Fitur tanggal dan waktu asli didekomposisi untuk mengekstrak informasi kontekstual. Fitur yang diciptakan antara lain fitur yang menangkap perbedaan pola perilaku antara hari kerja dan akhir pekan. Selain itu, perbedaan pola juga didasari dengan representasi siklus 24 jam. Hal ini memungkinkan model untuk memahami pola waktu tanpa terpengaruh oleh diskontinuitas waktu di belahan negara tertentu.

2. Agregat alamat IP

Penelitian ini menciptakan fitur baru dengan mengkorelasikan fitur IP Address dengan ID pelanggan. Fitur ini menghitung jumlah pelanggan unik yang menggunakan satu alamat IP yang sama. Nilai yang

tinggi pada fitur ini merupakan sinyal kuat adanya penggunaan proxy, VPN, atau serangan terkoordinasi (botnet).

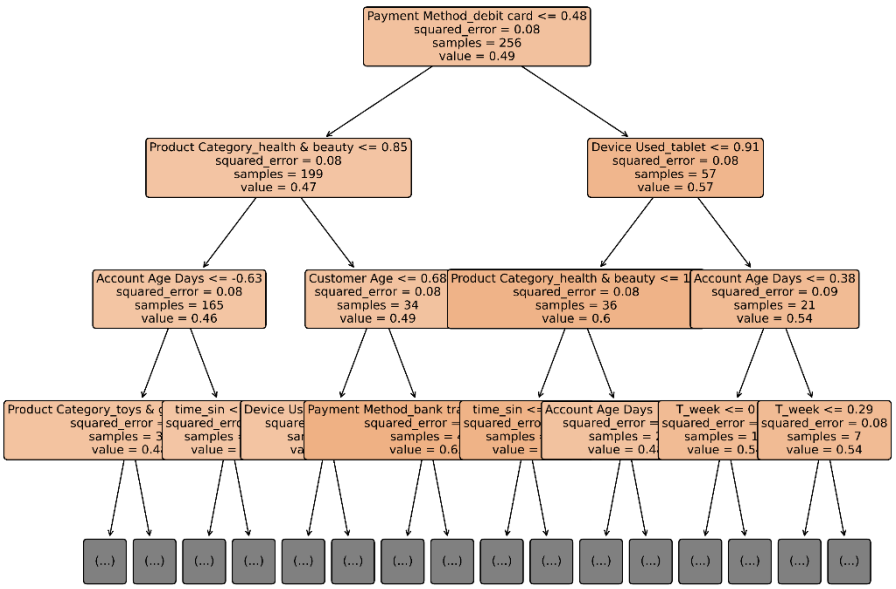
2.2.3 Pra-Pemodelan

Model machine learning beroperasi menggunakan data numerik, sehingga semua fitur kategorikal perlu ditransformasi. Penelitian ini menerapkan teknik One-Hot Encoding karena kemampuannya mengubah fitur kategorikal menjadi representasi numerik tanpa menciptakan urutan artifisial yang tidak logis. Dataset yang dihasilkan dari tahap ini memiliki fitur-fitur yang sepenuhnya numerik namun dengan skala yang heterogen. Heterogenitas skala ini berpotensi menyebabkan bias, di mana model akan lebih dipengaruhi oleh fitur dengan rentang nilai yang besar. Heterogenitas skala ini perlu diseragamkan melalui proses standarisasi data untuk mencegah model memberikan bobot yang tidak proporsional pada fitur dengan rentang numerik yang besar. Metode RobustScaler diterapkan karena kemampuannya yang superior dalam menangani pencilan ekstrem (outliers) yang umum ditemukan dalam data transaksi keuangan.

2.3 Pemodelan: Ensemble Learning

Penelitian ini memilih Ensemble Learning sebagai pendekatan pemodelan utama. Metode ini dipilih karena kemampuannya dalam mengurangi overfitting, dan menangkap hubungan non-linear yang kompleks antar fitur. Secara spesifik, algoritma yang digunakan adalah Isolation Forest, sebuah metode ensemble yang termasuk dalam kategori Unsupervised Learning. Dalam penelitian ini, Isolation Forest tidak dilatih menggunakan label Is Fraudulent, metode ini mampu mengidentifikasi anomali murni berdasarkan struktur dan kepadatan data. Ini memungkinkannya untuk menemukan pola-pola aneh yang mungkin tidak disadari atau terlewatkan oleh pengembang data. Metode ini secara inheren bekerja dengan cara mengisolasi titik data yang paling berbeda. Logika ini sangat cocok untuk kasus deteksi fraud, di mana aktivitas penipuan seringkali merupakan anomali yang langka di tengah lautan transaksi sah. Gambar 2 menyajikan visual dari salah satu decision tree yang dibangun dalam model Isolation Forest. Setiap node dalam pohon merepresentasikan suatu kondisi pemisahan berdasarkan nilai fitur tertentu (misalnya, Payment Method_debit card <= 0.48 atau Account Age Days <= -0.63).

Titik data akan melewati serangkaian pemisahan ini hingga mencapai daun. Dari visualisasi ini, terlihat bagaimana fitur-fitur yang telah direkayasa dan prapemrosesan (Payment Method, Product Category, Device Used, Account Age Days, time_sin, T_week) secara aktif digunakan oleh model untuk mengisolasi observasi. Node paling atas memisahkan transaksi berdasarkan Payment Method_debit card. Jalur ke kiri (true) dan ke kanan (false) memisahkan lagi berdasarkan Product Category_health & beauty atau Device Used_tablet. Anomali biasanya akan mengikuti jalur yang sangat pendek, terisolasi dengan sedikit pemisahan. Visualisasi ini secara intuitif menunjukkan bagaimana model secara hirarkis membedakan pola transaksi, memungkinkan deteksi anomali yang efisien dan tidak bergantung pada label historis.



Gambar 2. Pohon Keputusan dari Ensemble Isolation Forest

2.4 Evaluasi Model

Evaluasi model dalam penelitian ini dilakukan dengan pengecekan silang terhadap label asli Is Fraudulent untuk mengukur relevansi temuan. Metrik utama yang digunakan adalah Precision dan Recall. Precision mengukur tingkat alarm palsu, sementara Recall mengukur daya tangkap model terhadap kasus fraud yang telah dilabelkan pengembang pada dataset. Analisis terhadap kedua metrik ini menjadi krusial. Tingkat Recall yang wajar menunjukkan bahwa pola yang ditemukan model penelitian ini relevan dengan aktivitas fraud yang sudah dikenal. Di sisi lain, tingkat Precision membuka ruang bahwa anomali yang terdeteksi namun tidak ada di label asli berpotensi menjadi pola serangan generasi baru yang terlewatkan oleh sistem pelabelan sebelumnya.

3 Hasil dan Pembahasan

3.1 Hasil

Bab ini menyajikan hasil analisis data dan pemodelan yang telah dilakukan sesuai dengan kerangka metodologi yang diuraikan pada bab sebelumnya. Analisis ini terbagi menjadi beberapa bagian utama. Pertama, penelitian ini akan menyajikan hasil dari model deteksi mencurigakan unsupervised Isolation Forest, yang menjadi inti dari pendekatan penemuan pola dalam penelitian ini. Kedua, penelitian ini melakukan validasi kuantitatif dan kualitatif terhadap temuan model untuk mengukur relevansi dan mengkarakterisasi pola yang teridentifikasi. Terakhir, penelitian ini akan menggunakan visualisasi data untuk memperkuat temuan dan menyajikan bukti visual dari pola serangan yang berhasil dikenali.

Pemilihan model utama dalam penelitian ini jatuh pada Isolation Forest, sebuah pendekatan ensemble learning yang bersifat unsupervised. Keputusan ini bersifat strategis dan selaras dengan tujuan utama penelitian, yaitu untuk mengidentifikasi pola serangan generasi baru tanpa terikat pada definisi fraud yang sudah ada di dalam label historis dataset. Model ini dilatih pada keseluruhan data fitur yang telah direkayasa dan diskalakan, memungkinkannya untuk mempelajari struktur data secara holistik. Prinsip kerja Isolation Forest didasarkan pada hipotesis bahwa aktivitas anomali memiliki karakteristik yang berbeda secara fundamental dari aktivitas normal. Dalam konteks ini, perilaku transaksional yang mencurigakan seperti impulsivitas yang ditunjukkan pelaku serangan akan menciptakan pola yang secara statistik terisolasi dari mayoritas transaksi yang valid. Untuk mengkalibrasi sensitivitas model dalam mengenali pola-pola ini, penelitian ini menyoroti peran krusial dari hyperparameter contamination. Parameter ini berfungsi sebagai estimasi awal mengenai proporsi anomali di dalam dataset. Sebuah analisis sensitivitas kemudian dilakukan dengan menguji beberapa nilai contamination yang berbeda. Eksperimen ini bertujuan untuk secara kuantitatif mengetahui bagaimana perubahan parameter tersebut mempengaruhi trade-off antara kemampuan model dalam menangkap kasus fraud (Recall) dan akurasi dari anomali yang ditandai (Precision), sehingga dapat ditemukan titik keseimbangan yang paling informatif untuk analisis.

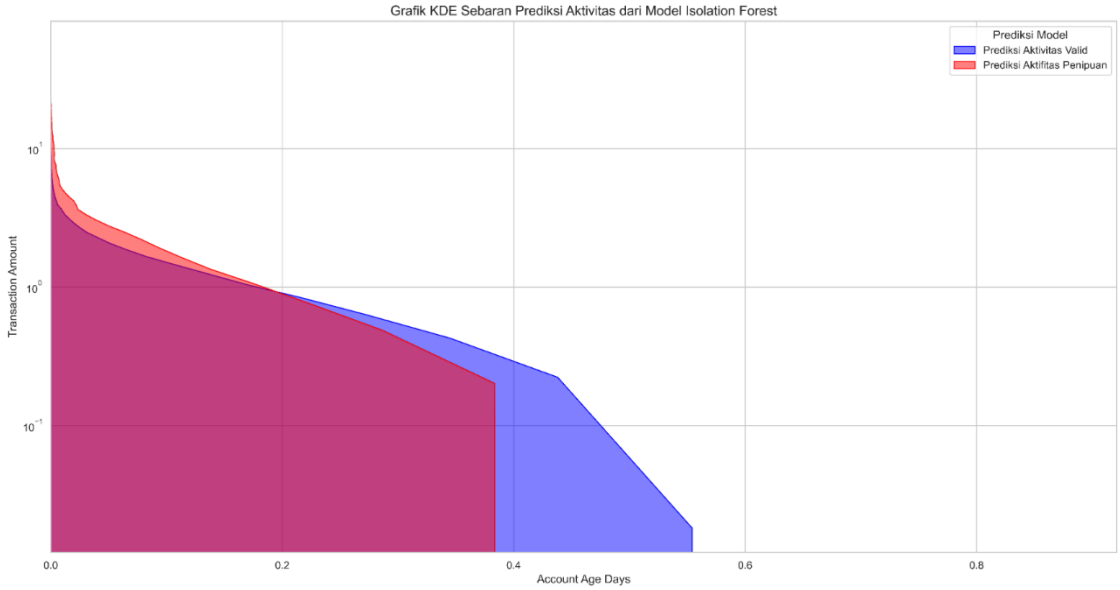
Table 1. Hasil Evaluasi Model Isolation Forest pada Berbagai Tingkat

Cont	Recall		Precision		F1-Score	
	1	0	1	0	1	0
0.01	0,03	0,99	0,13	0,95	0,04	0,97
0.05	0,11	0,95	0,11	0,95	0,11	0,95
0.1	0,19	0,9	0,1	0,95	0,13	0,93
0.2	0,32	0,81	0,08	0,96	0,14	0,88
0.3	0,43	0,71	0,07	0,96	0,12	0,81
0.4	0,53	0,61	0,07	0,96	0,12	0,74
0.5	0,62	0,51	0,06	0,96	0,11	0,66

Berdasarkan eksperimen yang penelitian ini kaji dihasilkan bahwa contamination 0.1 memberikan keseimbangan terbaik antara daya jangkau deteksi (Recall) dan akurasi peringatan (Precision), yang menghasilkan F1-score tertinggi di antara konfigurasi yang diuji. Model berhasil mencapai Recall sebesar 0.18 untuk kelas fraud. Ini membuktikan bahwa dari seluruh kasus fraud yang telah dilabelkan, model penelitian ini mampu secara mandiri menemukan kembali 18% di antaranya hanya dengan menganalisis struktur data. Ini memberikan validitas yang kuat bahwa pola mencurigakan yang dideteksi oleh model memang sangat relevan dengan aktivitas yang juga terlabel sebagai aktivitas penipuan. Kemampuan untuk menjaring lebih dari 13.000 kasus fraud yang terbukti tanpa pelatihan berlabel menunjukkan nilai strategis tersebut sebagai sistem peringatan dini. Di sisi lain, Precision sebesar 0.09 membuka pintu menuju argumen utama penelitian ini. Angka ini berarti bahwa dari seluruh transaksi yang ditandai sebagai aktivitas mencurigakan oleh model, hanya 9% yang cocok dengan label fraud yang ada. Ini menyisakan 91% aktivitas mencurigakan lainnya yang saat ini masih terlabel sebagai aktivitas sah. Berdasarkan eksperimen yang dilakukan terdapat potensi terbesar untuk menemukan pola serangan generasi baru yang belum teridentifikasi, yaitu transaksi-transaksi yang berhasil lolos dari sistem deteksi sebelumnya.

Inti dari kontribusi penelitian ini terletak pada analisis mendalam terhadap anomali yang ditemukan, terutama yang tidak teridentifikasi oleh label asli. Dengan melakukan studi kasus pada transaksi-transaksi yang diberi skor anomali tertinggi oleh model namun berlabel sah, penelitian ini berhasil mengidentifikasi sebuah pola serangan yang konsisten. Pola ini tidak bergantung pada satu indikator tunggal yang ekstrem, melainkan pada kombinasi beberapa sinyal risiko tingkat menengah yang muncul secara bersamaan. Indikasi yang pertama sebagai penguatan penemuan pola serangan penipuan yang baru yaitu pada transaksi yang berasal dari alamat IP yang juga digunakan oleh beberapa akun pelanggan lain dalam periode waktu yang berdekatan. Ini adalah indikasi kuat penggunaan layanan anonimisasi (VPN/Proxy) atau aktivitas botnet terkoordinasi. Analisa dilanjutkan dengan kategorisasi usia akun transaksi yang seringkali dilakukan oleh akun yang baru dibuat. Pelaku fraud cenderung menggunakan akun sekali pakai untuk menghindari pelacakan historis. Sebuah transaksi yang hanya memiliki salah satu dari ciri ini mungkin tidak akan terdeteksi oleh sistem berbasis aturan. Namun, model Isolation Forest, dengan kemampuannya melihat data secara holistik, berhasil mengidentifikasi bahwa kemunculan simultan dari beberapa karakteristik ini menciptakan sebuah aktivitas mencurigakan yang sangat kuat. Kemampuan untuk mengenali konvergensi risiko inilah yang penelitian ini definisikan sebagai pengenalan pola serangan generasi baru.

Visualisasi intuitif untuk validasi temuan dalam penelitian ini dibangun melalui penerapan teknik Estimasi Kepadatan Kernel (KDE) yang ditampilkan pada gambar 3.



Gambar 3. Visualisasi Grafik KDE

KDE mampu memetakan peta kepadatan data yang halus. Visualisasi ini secara efektif mengenali pusat gravitasi dari pola transaksi yang berbeda. Penyusunan visualisasi ini didasarkan pada dua fitur paling prediktif yang telah diidentifikasi melalui analisis kepentingan fitur, yaitu Account Age Days (usia akun dalam hari) dan Transaction Amount (nilai transaksi). Pemilihan kedua fitur ini bersifat strategis karena keduanya merepresentasikan dimensi perilaku yang fundamental dalam transaksi digital. Usia akun merepresentasikan tingkat kepercayaan historis seorang pelanggan, sementara nilai transaksi merepresentasikan besaran risiko pada satu titik waktu. Korelasi antara kedua fitur ini menjadi sangat krusial dalam membedakan antara aktivitas yang sah dan aktivitas penipuan. Pelanggan yang benar-benar sah umumnya menunjukkan pola maturasi, di mana aktivitas dibangun seiring berjalannya waktu yang tercermin pada usia akun. Di sisi lain, pelaku penipuan seringkali tidak memiliki kemewahan waktu sehingga harus memaksimalkan keuntungan secepat mungkin sebelum terdeteksi.

Visualisasi KDE yang dihasilkan secara efektif menyoroti adanya perbedaan fundamental antara distribusi perilaku transaksi normal dan mencurigakan. Terlihat jelas bahwa pusat kepadatan untuk aktivitas normal terkonsentrasi pada akun dengan usia yang lebih matang dan rentang nilai transaksi yang wajar. Sebaliknya, pola serangan yang berhasil diidentifikasi oleh model terkonsentrasi pada area spesifik, yaitu aktivitas transaksi bernilai tinggi yang dilakukan oleh akun yang usianya masih sangat baru. Pola ini mencerminkan pelaku sengaja membuat akun baru yang seringkali tanpa melalui proses verifikasi yang lengkap

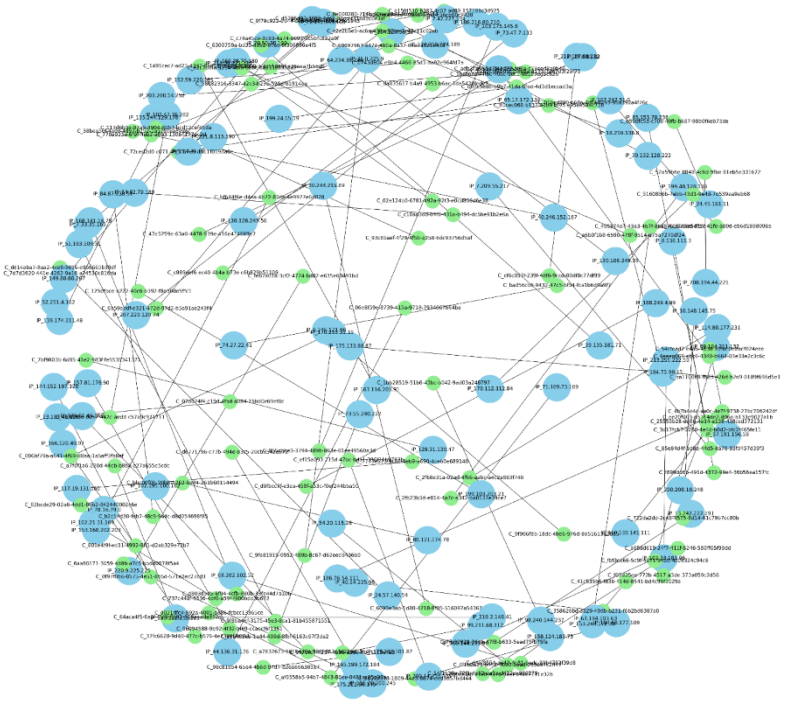
untuk kemudian melakukan transaksi sebesar mungkin dan menghilang. Dengan demikian, kemampuan untuk mengenali konsentrasi aktivitas anomali di pojok kuadran ini berfungsi sebagai sistem peringatan dini yang sangat kuat. Ini menjadi bukti visual bahwa pendekatan unsupervised yang penelitian ini usulkan tidak hanya menemukan anomali secara acak, tetapi secara spesifik berhasil mengenali dan memetakan sebuah pola serangan yang logis dan konsisten dengan taktik penipuan di dunia nyata.

Untuk memperdalam analisis Pola Serangan yang ditemukan, penelitian ini melakukan visualisasi grafik jaringan untuk memetakan hubungan antara dua entitas kunci: Akun Pelanggan dan Alamat IP. Seperti yang ditunjukkan pada Gambar 4, visualisasi ini terdiri dari dua jenis node: node biru (besar) yang merepresentasikan Alamat IP dan node hijau (kecil) yang merepresentasikan Akun Pelanggan (Customer ID). Sebuah garis (edge) ditarik jika sebuah akun melakukan transaksi menggunakan Alamat IP tersebut.

Visualisasi ini memberikan bukti kualitatif yang kuat atas temuan kuantitatif penelitian ini. Jika perilaku normal adalah satu akun terhubung ke satu atau beberapa IP, Gambar 3 dengan jelas menunjukkan pola anomali yang masif.

Terlihat jelas adanya beberapa supernode Alamat IP (node biru), yaitu di kuadran kanan bawah dan kiri atas, di mana satu Alamat IP tunggal digunakan oleh puluhan hingga ratusan Akun Pelanggan yang berbeda.

Temuan ini secara visual mengkonfirmasi hipotesis dari fitur rekayasa yang penelitian ini bangun. Pola yang tercipta ini adalah indikator kuat dari aktivitas terkoordinasi (seperti botnet) atau penggunaan layanan anonimisasi (VPN/Proxy) yang jamak digunakan pelaku fraud untuk menyamarkan jejak pelaku. Ini adalah bukti visual dari pola serangan yang berhasil diidentifikasi oleh model unsupervised, di mana risiko tidak terletak pada satu transaksi, melainkan pada relasi tersembunyi antar transaksi.



Gambar 4. Visualisasi Grafik Jaringan untuk Relasi Alamat IP dan Akun Pelanggan

3.2 Pembahasan

Bab ini membahas implikasi dari temuan penelitian dengan membandingkannya terhadap kerangka penelitian yang lebih luas. Diskusi difokuskan pada tiga pilar utama: validitas pendekatan unsupervised dalam konteks deteksi novelty, justifikasi pemilihan Isolation Forest dibandingkan metode anomali lainnya, dan signifikansi temuan pola serangan bentuk serangan terkoordinasi.

Hasil penelitian ini menunjukkan bahwa Isolation Forest (iForest) mampu mengidentifikasi 18% dari kasus fraud yang telah dilabelkan. Penelitian oleh Ahmed et al. [16] yang menggunakan hybrid-sampling ensemble model (model supervised) pada data fraud yang sangat tidak seimbang berhasil mencapai F1-score sebesar 91.5%. Keunggulan akurasi model supervised ini sangat bergantung pada ketersediaan data historis yang telah dilabeli, yang berarti model tersebut hanya dilatih untuk

mengenali pola fraud yang telah diketahui. Keterbatasan inilah yang menjadi celah (gap) utama. Temuan 18% recall dalam penelitian ini justru harus dimaknai sebagai validasi positif, bahwa sebuah model unsupervised murni, dengan tanpa label, mampu secara mandiri menemukan pola yang dianggap fraud oleh sistem konvensional. Lebih penting lagi, 91% anomali temuan model penelitian ini yang tidak ada di label asli adalah bukti keberhasilan dalam menemukan novelty. Ini sejalan dengan argumen bahwa sistem deteksi modern harus beralih dari sekadar prediksi reaktif ke penemuan ancaman (threat discovery) proaktif.

Pemilihan Isolation Forest sebagai inti metodologi didasarkan pada efisiensi dan paradigmanya yang unik. Untuk memvalidasi pilihan ini, kami membandingkannya dengan penelitian komparatif. Herreros et al. [17] melakukan studi komparasi antara K-Means, Local Outlier Factor (LOF), dan Isolation Forest untuk deteksi anomali pada transaksi kartu kredit. Hasilnya menunjukkan bahwa meskipun LOF memiliki akurasi yang bersaing, Isolation Forest secara signifikan lebih unggul dalam hal efisiensi komputasi dan kecepatan pemrosesan pada dataset bervolume besar. Efisiensi ini krusial untuk implementasi di dunia nyata pada ekosistem pembayaran digital yang memproses jutaan transaksi per detik. Paradigma iForest yang secara inheren mengisolasi anomali pada data normal seperti OCSVM, terbukti sangat efektif. Temuan pola serangan dalam penelitian kami, yang merupakan kombinasi multi-faktor, adalah jenis anomali kompleks yang dirancang untuk dideteksi oleh iForest, sementara mungkin terlewatkan oleh metode berbasis klaster seperti K-Means yang berfokus pada jarak sentroid.

Kontribusi utama penelitian ini adalah identifikasi pola serangan yang didorong oleh rekayasa fitur (terutama agregat Alamat IP dan usia akun). Pentingnya rekayasa fitur ini ditegaskan oleh Li et al. [18] yang berargumen bahwa behavior-based feature engineering (rekayasa fitur berbasis perilaku) seringkali lebih prediktif daripada data transaksi mentah. Fitur Agregat alamat IP pada penelitian ini adalah bentuk dari behavioral feature tersebut. Diskusi ini juga memperkuat saran untuk penelitian selanjutnya. Temuan anomali yang masif (91% dari deteksi) dapat disalurkan ke tahap kedua. Shanaa et al. [19] mengusulkan model hybrid di mana unsupervised learning (seperti iForest) bertindak sebagai penyaring (filter) untuk menghasilkan kandidat anomali, yang kemudian diumpankan ke model supervised (seperti XGBoost) untuk klasifikasi akhir. Pendekatan hybrid ini akan memanfaatkan kekuatan kedua dunia: kemampuan penemuan novelty dari iForest dan akurasi prediksi tinggi dari model supervised.

2.5. *Daftar Rujukan*

[1] [1] B. Indonesia, “Statistik Sistem Pembayaran dan Infrastruktur Pasar Keuangan (SPIP),” Bank Indonesia, 2024. https://www.bi.go.id/id/statistik/ekonomi-keuangan/spip/default.aspx (accessed Mar. 20, 2025).

[2] [2] Otoritas Jasa Keuangan, “POJK Nomor 21/2023 tentang Layanan Digital oleh Bank Umum.” 2023. [Online]. Available: https://ojk.go.id/id/regulasi/Pages/Layanan-Digital-oleh-Bank-Umum.aspx

[3] [3] Otoritas Jasa Keuangan, “Laporan Surveillance Perbankan Indonesia – Triwulan I 2023.” Jakarta, 2023. [Online]. Available: https://ojk.go.id/id/kanal/perbankan/data-dan-statistik/laporan-profil-industri-perbankan/Documents/Laporan Surveillance Perbankan Indonesia - Triwulan I 2023.pdf

[4] [4] N. Muhamad, “Mayoritas Masyarakat Tidak Yakin dengan Tingkat Keamanan Siber di Indonesia,” Databoks, 2023. https://databoks.katadata.co.id/layanan-konsumen-kesehatan/statistik/6777ef621af3ec4/mayoritas-masyarakat-tidak-yakin-dengan-tingkat-keamanan-siber-di-indonesia (accessed Mar. 19, 2025).

[5] [5] T. S. R. Institute, “Position Paper: Accelerating Inclusive and Fair Digital Transformation in Indonesia,” 2022. [Online]. Available: https://smeru.or.id/sites/default/files/publication/position_paper_accelerating_inclusive_and_fair_digital.pdf

[6] [6] F. B. of Investigation, “Internet Crime Report 2023,” Washington, D.C., 2023. [Online]. Available: https://www.ic3.gov/Media/PDF/AnnualReport/2023_IC3Report.pdf

[7] [7] A. Ali et al., “Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review,” Applied Sciences (Switzerland), vol. 12, no. 19. 2022. doi: 10.3390/app12199637.

[8] [8] H. Zhou, G. Sun, S. Fu, L. Wang, J. Hu, and Y. Gao, “Internet Financial Fraud Detection Based on a Distributed Big Data Approach with Node2vec,” IEEE Access, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3062467.

[9] [9] Y. Zhu, K. Xu, and W. Zhou, “Enhancing credit card fraud detection using hybrid neural network and SMOTE techniques,” J. Technol. Perform. Eng. Sci., vol. 3, no. 2, pp. 45–54, 2024, [Online]. Available: https://centuryscipub.com/index.php/jtpes/article/view/491

[10] [10] R. Li, Z. Liu, Y. Ma, D. Yang, and S. Sun, “Internet Financial Fraud Detection Based on Graph Learning,” IEEE Trans. Comput. Soc. Syst., vol. 10, no. 3, 2023, doi: 10.1109/TCSS.2022.3189368.

[11] [11] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-based anomaly detection,” ACM Trans. Knowl. Discov. Data, vol. 6, no. 1, p. 3, 2012, doi: 10.1145/2133360.2133363.

[12] [12] United Nations, “THE 17 GOALS | Sustainable Development.” 2025. [Online]. Available: https://sdgs.un.org/goals

[13] [13] S. Jagtap, “Fraudulent E-commerce Transactions,” Kaggle, 2023. https://www.kaggle.com/datasets/shriyashjagtap/fraudulent-e-commerce-transactions (accessed Mar. 29, 2025).

[14] [14] V. de Vargas, J. A. Schneider Aranda, R. dos Santos Costa, P. R. da Silva Pereira, and J. L. Victória Barbosa, “Imbalanced data preprocessing techniques for machine learning: a systematic mapping study,” Knowl. Inf. Syst., vol. 65, no. 1, pp. 31–57, 2023.

[15] [15] S. Galli, “Feature-engine: A Python package for feature engineering for machine learning,” J. Open Source Softw., vol. 6, no. 65, p. 3642, 2021.

[16] [16] K. H. Ahmed, S. Axelsson, Y. Li, and A. M. Sagheer, “A Credit Card Fraud Detection Approach Based on Ensemble Machine Learning Classifier with Hybrid Data Sampling,” Mach. Learn. with Appl., p. 100675, 2025.

[17] [17] A. Herreros-Mart\’\inez, R. Magdalena-Benedicto, J. Vila-Francés, A. J. Serrano-López, S. Pérez-D\’\iaz, and J. J. Mart\’\inez-Herráiz, “Applied Machine Learning to Anomaly Detection in Enterprise Purchase Processes: A Hybrid Approach Using Clustering and Isolation Forest,” Information, vol. 16, no. 3, p. 177, 2025.

[18] [18] E. Li, M. Chen, S. Xiang, and L. Chen, "Graph-Learning-Empowered Financial Fraud Detection: Progress and Future Directions," Intell. Comput..

[19] [19] M. Shanaa and S. Abdallah, "A hybrid anomaly detection framework combining supervised and unsupervised learning for credit card fraud detection," F1000Research, vol. 14, p. 664, 2025.

[20]