p-ISSN : 2406-7733
e-ISSN : 2597-9922

OPTIMALISASI DETEKSI KECURANGAN PADA TRANSAKSI *E-WALLET* MENGGUNAKAN ALGORITMA *ISOLATION FOREST* BERBASIS *BIG DATA*

Anggi Ferita Oktaviani Silalahi[1], Azhara Amelia. H[2], Sabrina Akva[3], Desni Paramitha Purba[4] , Fanny Ramadhani[5], Arnita[6]
Ilmu Komputer , Fakultas Matematika dan Ilmu Pengetahuan Alam , Universitas Negeri Medan
Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kab. Deli Serdang, Sumatera Utara
E-mail: * anggisilalahi338@gmail.com [1], azharaamelia14@gmail.com [2], sabrinaakva55110@gmail.com [3], purbadesnip@gmail.com [4], fannyr@unimed.ac.id[5], arnita@unimed.ac.id[6]

Abstrak - Kemajuan teknologi finansial telah mendorong adopsi layanan dompet digital (*e-wallet*) secara luas. Namun, peningkatan volume transaksi juga membawa risiko keamanan yang tinggi, khususnya terkait aktivitas kecurangan. Penelitian ini bertujuan membangun sistem deteksi kecurangan pada transaksi *e-wallet* menggunakan algoritma *Isolation Forest*, yang mampu mengidentifikasi anomali secara efisien tanpa memerlukan data berlabel. Dataset yang digunakan terdiri dari 6.362.620 transaksi *e-wallet* yang mencakup atribut numerik dan kategorikal. Proses penelitian meliputi tahapan preprocessing data, pelatihan model, dan evaluasi kinerja dengan *metrik precision*, *recall*, dan F1-*score*. Hasil evaluasi menunjukkan bahwa meskipun akurasi model mencapai 99%, *recall* terhadap transaksi fraud masih rendah, yaitu sebesar 4%, menandakan perlunya pendekatan lanjutan untuk meningkatkan sensitivitas model. Penelitian ini menunjukkan potensi *Isolation Forest* dalam mendeteksi pola transaksi anomali pada data berukuran besar serta memberikan dasar untuk pengembangan sistem keamanan finansial berbasis data.

Kata kunci: *anomaly detection,* deteksi kecurangan, e-wallet, Isolation Forest, machine learning.

I. PENDAHULUAN

Dalam beberapa tahun terakhir, teknologi finansial (*fintech*) telah mengalami pertumbuhan yang sangat pesat, dengan layanan dompet digital (*e-wallet*) menjadi salah satu sektor yang mengalami perkembangan signifikan. Kemudahan, kecepatan, dan kenyamanan yang ditawarkan *e-wallet* telah mendorong peningkatan penggunaannya secara masif di berbagai kalangan masyarakat. Menurut data dari Bank Indonesia, transaksi *e-wallet* mencapai lebih dari 1 miliar kali dalam satu tahun dengan nilai transaksi yang terus meningkat setiap tahunnya. Namun, seiring dengan melonjaknya volume transaksi digital, muncul pula tantangan serius dalam aspek keamanan, terutama terkait praktik kecurangan (*fraud*) yang semakin kompleks dan sulit dikenali.

Kecurangan dalam transaksi *e-wallet* tidak selalu bersifat eksplisit dan dapat berupa aktivitas mencurigakan yang sulit dideteksi secara manual, seperti pengisian saldo (*top-up*) dalam jumlah besar secara tiba-tiba, transaksi dengan nilai tidak wajar pada waktu yang tidak lazim, atau aktivitas dari lokasi geografis yang tidak konsisten dengan kebiasaan pengguna. Tantangan utama dalam mendeteksi kecurangan ini terletak pada besarnya volume data serta kompleksitas pola transaksi yang sangat beragam dan tidak selalu mengikuti pola yang jelas. Oleh karena itu, deteksi kecurangan memerlukan pendekatan berbasis data mining dan *machine learning* yang mampu mengidentifikasi anomali atau penyimpangan dalam data secara efisien.

Salah satu algoritma yang relevan dan banyak digunakan untuk mendeteksi anomali dalam dataset besar adalah *Isolation Forest*. Algoritma ini bekerja dengan prinsip mengisolasi data poin yang berbeda secara struktural dari mayoritas data lainnya, dan dinilai sangat efisien dalam mengolah data yang tidak berlabel. Keunggulan algoritma ini terletak pada kemampuannya untuk mendeteksi penyimpangan tanpa memerlukan proses pelabelan data yang mahal dan rumit, menjadikannya sangat cocok untuk kasus fraud detection dalam sistem transaksi *e-wallet*.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan sistem deteksi kecurangan transaksi *e-wallet* berbasis data simulasi menggunakan algoritma *Isolation Forest*. Penelitian ini mencakup tahapan *preprocessing* data, penerapan model, serta evaluasi performa menggunakan *metrik precision*, *recall*, dan F1-*score*. Hasil yang diperoleh menunjukkan bahwa meskipun model mampu mencapai akurasi tinggi secara keseluruhan (99%), kemampuan dalam mendeteksi transaksi *fraud* masih terbatas, dengan *recall* untuk kelas *fraud* hanya mencapai 4%. Hal ini menjadi dasar penting untuk memberikan *insight* terhadap pola-pola transaksi anomali yang berhasil diidentifikasi oleh model dan membuka peluang pengembangan lebih lanjut terhadap metode deteksi *fraud* berbasis *machine learning*. Selain itu, penelitian ini menawarkan kontribusi berupa penerapan metode *unsupervised* pada kasus nyata dengan dataset yang sangat tidak seimbang, sebagai upaya untuk mendeteksi *fraud* secara lebih efisien dan adaptif.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

II. TINJAUAN PUSTAKA

Deteksi kecurangan (*fraud detection*) merupakan salah satu aspek krusial dalam menjaga integritas sistem keuangan modern, terutama dalam konteks transaksi digital yang semakin berkembang pesat. *Fraud* didefinisikan sebagai tindakan ilegal yang mengakibatkan kesalahan pelaporan keuangan atau penyalahgunaan sumber daya organisasi secara sengaja untuk memperoleh keuntungan pribadi. Menurut Elisabeth dan Simanjuntak (2020), *fraud* termasuk tindakan melawan hukum yang menuntut perhatian serius dari auditor untuk dapat dideteksi dan dicegah secara efektif. Dalam konteks hukum, *fraud* dalam transaksi digital juga dikategorikan sebagai tindak pidana yang berpotensi menjadi bagian dari kejahatan terorganisir dan pencucian uang (Pantow, 2025). Sistem informasi keuangan seperti *e-wallet*, *fraud* dapat muncul dalam bentuk transaksi palsu, penyalahgunaan akun, hingga eksploitasi celah keamanan sistem. Permasalahan utama dalam mendeteksi *fraud* adalah karakteristiknya yang tersembunyi dan jarang terjadi, menjadikannya sulit dikenali dengan metode konvensional, terutama karena ketidakseimbangan distribusi kelas (*imbalance class problem*) sebagaimana dijelaskan oleh Phua (2021).

Transaksi *e-wallet* sendiri merupakan bagian dari sistem pembayaran digital yang memungkinkan pengguna melakukan transaksi secara online dengan lebih mudah dan cepat. Di Indonesia, penggunaan dompet digital meningkat secara signifikan dan menjadi alat pembayaran non-tunai yang populer dalam berbagai transaksi berbasis internet. Menurut Al Qardh et al. (2020), pertumbuhan sistem pembayaran digital di Indonesia sangat dipengaruhi oleh perkembangan teknologi dalam era Revolusi Industri 4.0. Sunarsa dan Fauzi (2023) menyatakan bahwa *e-wallet* menyimpan, memproses, dan mentransfer uang dalam bentuk informasi digital, di mana proses pemindahan dana diinisiasi melalui perangkat elektronik. Sistem ini mencakup berbagai komponen seperti aplikasi transfer, infrastruktur jaringan, serta peraturan dan prosedur penggunaan. Informasi yang terkandung dalam transaksi *e-wallet*, seperti nominal, waktu, lokasi, jenis transaksi, dan identitas pengguna, sangat kaya dan kompleks. Data ini memiliki potensi besar untuk digunakan dalam mengidentifikasi pola-pola abnormal yang mungkin mengindikasikan kecurangan (Khoironnisa, 2024).

Peran big data dalam konteks keamanan finansial juga tidak dapat diabaikan. Big data menggambarkan kumpulan data dalam jumlah besar dan beragam, baik yang terstruktur maupun tidak terstruktur, yang terus tumbuh secara cepat. Sawitri (2019) menyebutkan tiga karakteristik utama big data, yaitu *volume*, *variety*, dan *velocity*. Dalam sistem keuangan digital, pemanfaatan big data memungkinkan analisis transaksi dalam jumlah masif secara *real-time* untuk mendeteksi potensi anomali. Hashem et al. (dalam Jabbar et al., 2020) menunjukkan bahwa integrasi big data *analytics* dengan *machine learning* memberikan keunggulan signifikan dalam meningkatkan keamanan siber, khususnya untuk sistem keuangan berbasis digital. Analisis data secara cepat dan akurat menjadi krusial dalam mengenali indikasi *fraud* sebelum menimbulkan kerugian besar.

Salah satu metode *machine learning* yang banyak digunakan dalam deteksi anomali pada big data adalah algoritma *Isolation Forest*. Algoritma ini diperkenalkan oleh Liu et al. (dalam Ahmad Zulfikar, 2023) sebagai metode *unsupervised learning* berbasis pohon keputusan yang dirancang untuk mengisolasi data-data anomali secara efisien. Proses deteksi dalam *Isolation Forest* terbagi menjadi dua tahap utama, yaitu tahap pelatihan di mana pohon-pohon isolasi dibentuk dari sampel data, dan tahap evaluasi di mana setiap data point diberi skor anomali berdasarkan seberapa cepat ia terisolasi dalam pohon. Menurut Ahmad Zulfikar (2023), *Isolation Forest* memiliki sejumlah keunggulan, di antaranya tidak memerlukan data berlabel, efisien dalam mendeteksi outlier, serta mampu menangani data berdimensi tinggi dengan kompleksitas waktu yang linear terhadap ukuran data. Studi lain oleh Wibawa dan Karyawati (2023) juga menunjukkan bahwa algoritma ini efektif dalam mendeteksi anomali pada data transaksi keuangan digital melalui pendekatan *exploratory data analysis.* Selain itu, algoritma lain seperti *Naive Bayes* dan *Decision Tree* juga telah diteliti untuk digunakan dalam deteksi anomali lalu lintas jaringan dan transaksi digital, meskipun *Isolation Forest* menunjukkan kinerja yang unggul dalam banyak kasus (Mendrofa et al., 2025).

Berbagai algoritma *machine learning* telah diterapkan dalam konteks deteksi penipuan di sektor keuangan, sebagaimana dibuktikan oleh Mustika et al. (2021) melalui studi kasus pada perusahaan pembiayaan konsumen ritel yang menunjukkan efektivitas pendekatan ini dalam mengidentifikasi pola *fraud* dari data *historis* transaksi. Dengan demikian, algoritma ini menjadi salah satu pendekatan yang menjanjikan dalam sistem deteksi *fraud* yang skalabel dan adaptif terhadap dinamika data transaksi digital.

III. METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif dengan memanfaatkan algoritma unsupervised learning *Isolation Forest* untuk mendeteksi anomali dalam transaksi *e-wallet*. Metode ini dipilih karena kemampuannya dalam mengidentifikasi data outlier secara efisien, bahkan

p-ISSN : 2406-7733
e-ISSN : 2597-9922

dalam dataset berskala besar tanpa memerlukan pelabelan data secara manual.

Dataset yang digunakan terdiri dari 6.362.620 data transaksi *e-wallet* yang mencakup 11 atribut utama, di antaranya adalah jumlah transaksi, jenis transaksi, waktu, saldo awal dan saldo akhir dari pengirim dan penerima, serta label *isFraud*. Seluruh data dianalisis secara komputasional menggunakan bahasa pemrograman Python.

Proses penelitian dimulai dengan pemuatan dataset dan eksplorasi awal untuk memastikan kelengkapan data. Tidak ditemukan nilai kosong dalam dataset ini, sehingga data dapat langsung digunakan dalam tahap *preprocessing*. Selanjutnya dilakukan konversi fitur kategorikal seperti type menjadi bentuk numerik melalui Label *Encoding* agar dapat diproses oleh algoritma. Beberapa fitur numerik kemudian dinormalisasi untuk menyamakan skala dan menghindari dominasi nilai-nilai besar dalam proses pembelajaran.

Model *Isolation Forest* dilatih dengan konfigurasi awal berupa *n_estimators* = 150 dan *contamination* = 0.05, serta *random_state* = 42 untuk reprodusibilitas. Algoritma ini bekerja dengan cara membangun pohon secara acak untuk memisahkan data dan mengukur seberapa cepat suatu titik data dapat diisolasi, yang mencerminkan tingkat keanehan (anomali). Model dijalankan secara paralel untuk mempercepat komputasi terhadap jutaan data. Evaluasi model dilakukan secara kuantitatif dengan memanfaatkan label *ground truth* yang tersedia (*isFraud*), meskipun model dilatih tanpa label. Metrik yang digunakan meliputi *precision, recall*, dan F1-*score* untuk menilai efektivitas deteksi anomali, khususnya dalam kondisi ketidakseimbangan kelas. Nilai evaluasi ini memberikan gambaran objektif terhadap kinerja model dalam membedakan transaksi normal dan anomali.

IV. HASIL DAN PEMBAHASAN

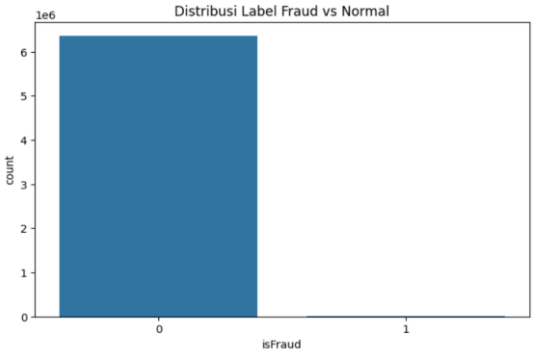
Dataset yang digunakan dalam penelitian ini terdiri dari 6.362.620 entri transaksi dengan 11 atribut yang mencakup informasi terkait jenis transaksi, jumlah, nama pengirim dan penerima, saldo sebelum dan sesudah transaksi, serta status transaksi apakah merupakan penipuan (*isFraud*) atau tidak. Hasil eksplorasi awal menunjukkan bahwa tidak terdapat nilai kosong (*missing values*) pada seluruh kolom, sehingga tidak diperlukan tahapan imputasi data.

Analisis statistik deskriptif dilakukan terhadap fitur *amount* untuk mengidentifikasi karakteristik transaksi normal dan *fraud*. Hasil analisis menunjukkan perbedaan yang signifikan antara kedua kelas:

Tabel 1. Perbandingan Statistik Rata Rata Transaksi

prediksi	mean	Std deviasi	median	maksimum
0	178.197	596.237	74.684	92.445.516
1	1.467.967	2.404.253	441.423	10.000.000

Transaksi *fraud* cenderung memiliki nilai nominal yang jauh lebih besar dibandingkan transaksi normal. Hal ini menjadi indikasi awal bahwa jumlah transaksi merupakan fitur penting dalam deteksi penipuan.



Gambar 1. Distribusi Label Transaksi *Fraud* dan *Non-Fraud*

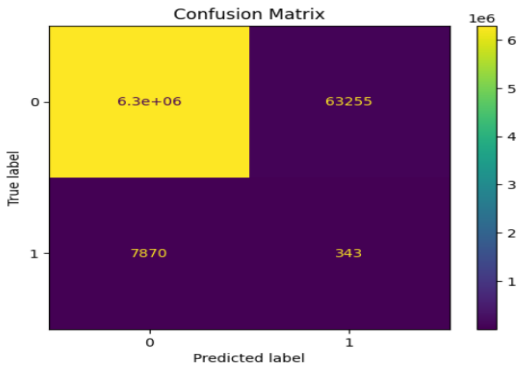
Pemrosesan Data numerik dinormalisasi menggunakan teknik standardisasi untuk menyamakan skala antar fitur. Proses ini bertujuan agar algoritma dapat mendeteksi anomali secara optimal tanpa terpengaruh perbedaan skala antar fitur. Contoh hasil normalisasi menunjukkan distribusi nilai yang sudah terpusat di sekitar nol.

Algoritma *Isolation Forest* digunakan karena kemampuannya dalam mendeteksi outlier pada dataset berskala besar tanpa membutuhkan label yang seimbang. Proses pelatihan model dilakukan dengan 150 *estimator*, dan berhasil dijalankan dengan waktu komputasi relatif cepat meskipun ukuran data sangat besar. Evaluasi dilakukan menggunakan *classification report* yang meliputi *metrik precision*, *recall*, f1-*score*, dan akurasi.

Tabel 2. Hasil Klasifikasi Model *Isolation Forest*

Label	Presisi	Recall	Skor F1	Jumlah Data
0	1	0,99	0,99	6.354.407
1	0,01	0,04	0,01	8.123
Akurasi			0,99	6.362.620
Rata rata Makro	0,50	0,52	0,50	-
Rata rata Terimbang	1	0,99	0,99	

p-ISSN : 2406-7733
e-ISSN : 2597-9922



Gambar 2. *Confusion matrix* hasil klasifikasi model *Isolation Forest*.

Analisis lanjutan dilakukan untuk memahami karakteristik transaksi yang terdeteksi sebagai *fraud*. Rata-rata nilai pada fitur-fitur utama menunjukkan bahwa transaksi fraud memiliki nominal saldo yang lebih tinggi pada akun pengirim dan penerima dibandingkan transaksi normal.
Hasil ini memperkuat indikasi bahwa nominal besar merupakan ciri khas dari transaksi penipuan.

V. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini menunjukkan bahwa algoritma *Isolation Forest* mampu memproses dataset transaksi *e-wallet* dalam skala besar secara efisien, dengan tingkat akurasi keseluruhan mencapai 99%. Meskipun demikian, hasil tersebut belum sepenuhnya merepresentasikan kemampuan model dalam mendeteksi transaksi penipuan, mengingat ketidakseimbangan kelas yang sangat tinggi dalam data. Transaksi *fraud* hanya mencakup sebagian kecil dari keseluruhan data, namun memiliki karakteristik yang berbeda secara signifikan, terutama dari sisi jumlah nominal transaksi dan saldo akun yang terlibat. Sayangnya, model hanya mampu mengenali sebagian kecil dari transaksi *fraud* dengan nilai *recall* sebesar 4% dan F1-*score* sebesar 1%, yang mengindikasikan rendahnya sensitivitas model terhadap kelas minoritas.

Saran

Ketidakseimbangan distribusi label antara transaksi normal dan *fraud* menjadi tantangan utama dalam pengembangan model deteksi anomali yang akurat. Oleh karena itu, dibutuhkan pendekatan lanjutan untuk meningkatkan performa deteksi, seperti Oversampling pada kelas *fraud*, seperti dengan *SMOTE*, Penggunaan model berbasis supervised learning, misalnya *Random Forest* atau *XGBoost* dengan penyesuaian *threshold*, serta Penyusunan fitur tambahan yang dapat menangkap pola perilaku *fraud* secara lebih spesifik. Dengan perbaikan ini, sistem deteksi *fraud* berbasis *machine learning* diharapkan mampu memberikan hasil yang lebih baik dan mendukung keamanan transaksi pada *platform e-wallet* secara lebih optimal.

DAFTAR PUSTAKA

Ahmad Zulfikar, F. A. R. N. A. (2023). Deteksi anomali menggunakan Isolation Forest belanja barang persediaan konsumsi pada satuan kerja Kepolisian Republik Indonesia. *Jurnal Manajemen Perbendaharaan*, *4*(1), 1–15.

Al Qardh, J., Tarantang, J., Awwaliyah, A., Astuti, M., & Munawaroh, M. (2020). *Perkembangan sistem pembayaran digital pada era revolusi industri 4.0 di Indonesia*. IAIN Palangka Raya.

Elisabeth, D. M., & Simanjuntak, W. A. (2020). Analisis review pendeteksian kecurangan (fraud). *Methosika: Jurnal Akuntansi dan Keuangan Methodist*, *4*(1), 9–18.

Jabbar, A., Akhtar, P., & Dani, S. (2020). *Real-time big data processing for instantaneous marketing decisions: A problematization approach.* Industrial Marketing Management, *90*, 558–569.

Khoironnisa. (2024). *Perlindungan terhadap pengguna e-wallet aetas hilangnya saldo pada aplikasi DANA dalam sistem pembayaran digital*. *Skripsi*. Universitas Islam Negeri Syarif Hidayatullah Jakarta.

Mendrofa, M. J., Las, K. J. D., & Budiati, H. (2025). Deteksi anomali traffic pada jaringan komputer menggunakan Naive Bayes, Decision Tree, dan Isolation Forest. *Jurnal Informatika dan Ilmu Komputer*, *5*(1), 45–56.

Mustika, N. I., Nenda, B., & Ramadhan, D. (2021). Algoritma machine learning dalam deteksi penipuan: Studi kasus pada perusahaan pembiayaan konsumen ritel. *Asia Pacific Fraud Journal*, *6*(2), 213–221.

Pantow, R. T. (2025). Tindak pidana penipuan dalam transaksi online sebagai kejahatan terorganisir dan kaitannya dengan pencucian uang. *Jurnal Hukum dan Kriminologi*, *7*(2), 112–125.

Phua, C. et al. (2021). A Comprehensive Survey of Data Mining-Based Fraud Detection Research. *Artificial Intelligence Review, 55,* 1985-2033

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Sawitri, D. (2019). Revolusi Industri 4.0: Big Data Menjawab Tantangan Revolusi Industri 4.0. *Jurnal Ilmiah Maksitek, 4*(3), 1-9.

Sunarsa, S., & Fauzi, I. N. (2023). Tinjauan hukum Islam tentang mekanisme transaksi e-wallet. *Jurnal Hukum Ekonomi Syariah (JHESY)*, *2*(1), 226–238.

Wibawa, I. M. S. T., & Karyawati, A. A. I. N. E. (2023). Isolation Forest dengan exploratory data analysis pada anomaly detection untuk data transaksi. *Jurnal Nasional Teknologi Informasi dan Aplikasnya*, *1*(3), 803–810.