ANALISIS KECURANGAN DALAM MENGHADAPI PENIPUAN DI SITUS *E-COMMERCE* MENGGUNAKAN *RANDOM FOREST* ; PENDEKATAN *MACHINE LEARNING* BERBASIS AI

[1]Ummi Kolbia, [2]Nova Dahliyanti,*
[1,2]Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia
[1]miikolbia@gmail.com, [2]a14893nov@gmail.com
**) Penulis Korespondensi*

Abstrak

Di era digital yang sedang berkembang pesat ini, fenomena e-commerce telah menjadi sorotan utama. Pesatnya pertumbuhan e-commerce telah menarik semakin banyak pengguna. Namun, kasus penipuan yang canggih dan dinamis meningkat seiring meningkatnya volume transaksi. Fenomena ini tidak hanya menimbulkan risiko kerugian finansial bagi pembeli dan penjual, tetapi juga mengancam kepercayaan yang sangat penting dalam industri e-commerce. Untuk mengatasi masalah ini, penulis menggunakan Random Forest pendekatan Machine Learning berbasis AI dalam menganalisis dan menemukan pola-pola kecurangan untuk menghadapi penipuan di situs e-commerce. Dengan fokus pada data transaksi e-commerce, penelitian ini menganalisis berbagai fitur transaksi untuk mengidentifikasi pola-pola yang berkaitan dengan aktivitas penipuan. Metodologi penelitian melibatkan pengumpulan dataset komprehensif dari transaksi e-commerce, pre-processing data, seleksi fitur, dan implementasi model Random Forest . Model ini dilatih dan divalidasi menggunakan teknik cross-validation untuk memastikan keandalan dan generalisasi. Performa model dievaluasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Random Forest yang dikembangkan mampu mendeteksi penipuan dengan tingkat akurasi 86% dan presisi 86%. Temuan ini memberikan wawasan berharga bagi pengembangan sistem keamanan e-commerce yang lebih robust dan adaptif terhadap ancaman penipuan yang terus berkembang.

***Kata Kunci**: Penipuan, e-commerce, random forest , machine learning.*

Abstract

In this rapidly expanding digital age, the phenomenon of e-commerce has become a major hit. The rapid growth of e-commerce has attracted more and more users. However, sophisticated and dynamic cases of fraud increased as the volume of transactions increased. This phenomenon not only endangers financial losses to buyers and sellers, but also threatens a very important trust in the e-commerce industry. To address this problem, the writer USES Random Forest , the ai machine learning approach, in analyzing and discovering fraudulent patterns to confront fraud at the site e-commerce. Focusing on trade data from e-commerce, the study analyzes various transaction features to identify patterns related to fraudulent activities. The research methodology involves the comprehensive datasset-gathering of e-commerce transactions, pre-processing data, selection features, and implementation of the Random Forest model. It is trained and validated using cross-validation techniques to ensure reliability and generalization. Model performance is evaluated using metrics such as accuracy, precision, recall, and f1-score. Research shows that the Random Forest model developed is able to detect fraud at an rate of 86% accuracy and 86% precision. These findings provide valuable insights into the developing security systems of e-commerce's more robust and adaptive to the growing threats of fraud.

***Kewords**: Fraud, e-commerce, random forest , machine learning.*

PENDAHULUAN

Di era digital seperti sekarang ini, *e-commerce* telah menjadi salah satu pilar utama perekonomian global, termasuk di Indonesia. Pertumbuhan *e-commerce* membawa banyak manfaat seperti kemudahan berbelanja, akses produk yang lebih luas, dan efisiensi transaksi. Menurut Statista Market Insights, pengguna *e-commerce* di Indonesia mencapai 178,94 juta pada tahun 2022 dan diproyeksikan mencapai 196,47 juta pada akhir tahun 2023[1]. Kementerian Perdagangan (Kemendag) memperkirakan nilai transaksi perdagangan digital atau *e-commerce* mencapai Rp533 triliun pada 2023 dibandingkan tahun sebelumnya yang tercatat Rp476 triliun[2].

Namun dibalik pertumbuhan tersebut, terdapat beberapa tantangan dan permasalahan yang perlu disikapi secara serius. Salah satu masalah utama yang muncul adalah meningkatnya kasus penipuan dalam transaksi *e-commerce*. Beberapa Platform *e-commerce* besar di Indonesia, seperti Tokopedia, Bukalapak, Shopee, Lazada, Blibi, telah mengalami kasus penipuan yang signifikan, diantaranya penipuan pembayaran, produk palsu atau tidak sesuai dengan yang dideskripsikan, dan pencurian identitas. Berdasarkan data survei SimilarWeb, Shopee mencatatkan pengunjung situs terbanyak dengan 235 miliar kunjungan pada Januari-Desember 2023[3]. Berbagai masalah yang telah dipaparkan di atas menunjukkan bahwa penipuan dalam *e-commerce* merupakan isu serius yang memerlukan solusi efektif. Penerapan model Machine Learning, khususnya Algoritma *Random Forest* , menjadi salah satu solusi yang dapat diandalkan untuk memprediksi dan mengidentifikasi pola-pola penipuan. Model ini dapat menganalisis berbagai variabel dan data transaksi untuk menentukan kemungkinan terjadinya penipuan berdasarakan hasil dari penelitian sebelumnya [4][5]. Teknologi ini memiliki potensi besar untuk mendeteksi pola-pola kecurangan secara lebih efektif dan efisien. Dengan menggunakan data transaksi, data pengguna, dan data historis kecurangan, model ini dapat dilatih untuk mengidentifikasi transaksi mencurigakan dan memberikan peringatan dini kepada pengguna serta Platform. Oleh karena itu, penulis tertarik untuk melakukan penelitian dengan judul "Analisis Kecurangan dalam Menghadapi Penipuan di Situs *E-commerce* Menggunakan *Random Forest* : Pendekatan Machine Learning Berbasis AI."

E-commerce

E-commerce adalah penggunaan internet dan web untuk transaksi bisnis. *e-commerce* juga adalah proses pembelian dan penjualan produk, jasa, dan informasi melalui jaringan komputer termasuk internet.[6]

E-commerce singkatan dari elektronik *commerce*, yang berarti sistem pemasaran yang dilakukan secara elektronik atau melalui media elektronik. *E-commerce* ini mencakup

pengiriman, penjualan, pembelian, promosi, dan layanan yang diberikan oleh produk melalui sistem elektronik seperti internet atau jenis jaringan komputer yang lain. [7]

Kecurangan di E-commerce

Analisis kecurangan adalah bidang yang mencakup banyak disiplin ilmu, termasuk ilmu komputer, statistik, dan bidang pengetahuan seperti keuangan atau asuransi, untuk menemukan aktivitas penipuan. Analisis jaringan sosial dan analisis sentimen digunakan untuk menemukan hubungan tersembunyi antara entitas. Penekanan utamanya adalah menemukan pola anomali yang mungkin tidak terlihat oleh manusia.[8]

Kecurangan dalam *e-commerce* dapat berupa tindakan ilegal yang dilakukan untuk menipu atau mendapatkan keuntungan secara tidak sah, seperti mengakses situs toko tanpa izin, menggunakan elemen situs tanpa izin, atau melanggar peraturan keamanan data.[9]

Penipuan di Situs E-commerce

Penipuan di situs *e-commerce* mengacu pada praktik curang atau menipu yang terjadi pada platform belanja *online*, seperti menyembunyikan informasi penting, menjual barang yang tidak sesuai dengan deskripsi, atau menipu dalam proses pembayaran.

Adapun faktor-faktor yang menyebabkan penipuan di situs *e-commerce* sebagai berikut;

a. Identitas palsu: penipu mungkin menggunakan identitas palsu atau mencuri identitas orang lain untuk melakukan transaksi.
b. Metode pembayaran: seperti transfer bank, kartu kredit, atau metode pembayaran online dapat digunakan untuk melakukan penipuan.
c. Kurangnya keamanan situs: situs *e-commerce* tidak memiliki sistem keamanan yang kuat, maka rentan terhadap serangan *hacker* atau pencurian data, yang dapat digunakan untuk melakukan penipuan.
d. Produk atau layanan palsu: Penjual yang tidak jujur dapat menjual barang atau layanan yang palsu atau tidak sesuai dengan deskripsi, memperdaya pelanggan.

Random Forest (RF)

Random Forest adalah salah satu metode *Machine Learning*, yang menggunakan beberapa *Decision Tree* dilatih secara individual dan membagi setiap atribut menjadi subset yang acak. *Random Forest* memiliki beberapa kelebihan, termasuk efisiensi penyimpanan data, ketahanan terhadap outliers, dan kemampuan untuk meningkatkan akurasi dalam kasus data yang hilang. Selain itu, *Random Forest* memiliki proses seleksi fitur yang dapat meningkatkan kinerja model klasifikasi. [10]

Berikut adalah kelebihan dari algoritma *Random Forest* :

a. Kuat terhadap data outlier (pencilan data).
b. Bekerja dengan baik dengan data non-linear.
c. Risiko overfitting lebih rendah.

d. Berjalan secara efisien pada kumpulan data yang besar.

e. Akurasi yang lebih baik daripada algoritma klasifikasi lainnya.

Machine Learning

Machine Learning adalah cabang artificial intelligence yang memungkinkan mesin belajar dari data atau pengalaman sebelumnya (data historis) untuk melakukan perintah tertentu. Terdapat tiga kategori algoritma dalam *Machine Learning*: pengajaran yang diawasi, pengajaran yang tidak diawasi, dan pengajaran pendukung. Setiap kategori memiliki prosedur pengajaran yang unik dan berbeda-beda. Proses belajar ini mencakup pengumpulan data, ekstraksi data, pemilihan model atau algoritma pembelajaran, pelatihan model atau algoritma tersebut, dan evaluasi hasil pembelajaran mesin. [11]

Adapun kelebihan *Machine Learning*, sebagai berikut :

a. Otomatisasi: Dapat melakukan hal-hal secara otomatis tanpa bantuan manusia. *Machine Learning* memiliki kemampuan untuk mempelajari pola tertentu, sehingga, seperti manusia, dapat berkembang tanpa pengkodean.

b. Peningkatan akurasi: Penggunaan data yang lebih besar dan kompleks membuat hasil lebih akurat. *Machine Learning* dapat membantu dalam berbagai industri seperti medis, keuangan, dan lain-lain karena dapat meningkatkan akurasi keputusan yang dibuat melalui proses pembelajaran data terlebih dahulu.

c. Penggunaan data yang banyak: Dapat menangani data yang sangat besar dan kompleks. *Machine Learning* sangat bermanfaat dalam berbagai industri seperti analisis data, pengembangan produk, dan lain-lain.

d. Mampu Meningkatkan Efisiensi: *Machine Learning* dapat meningkatkan efisiensi dalam berbagai tugas seperti pengolahan data, pengembangan produk, dan lain-lain, sehingga sangat berguna dalam berbagai industri seperti teknologi, keuangan, dan lain-lain.

Kecerdasan Buatan (AI)

Kecerdasan buatan (AI) adalah bidang ilmu komputer yang berfokus pada pembuatan agen cerdas, yaitu sistem yang dapat merasakan lingkungan mereka, belajar, membuat keputusan, dan mengambil tindakan untuk mencapai tujuan tertentu. AI bertujuan untuk membuat mesin yang dapat beradaptasi, berpikir, dan bertindak seperti manusia.[12]

Dalam hal keamanan digital, *artificial intelligence* (AI), atau yang diartikan sebagai kecerdasan buatan, adalah subjek yang sangat kontroversial. Selain itu, aplikasinya sangat luas, mulai dari menjaga privasi pengguna hingga menjaga integritas *platform* di seluruh dunia. Dalam era *e-commerce* yang berkembang pesat dimana transaksi lintas batas terjadi setiap detik, AI adalah alat penting untuk melindungi diri dari penipuan yang semakin canggih.

Deteksi Kecurangan dalam E-commerce

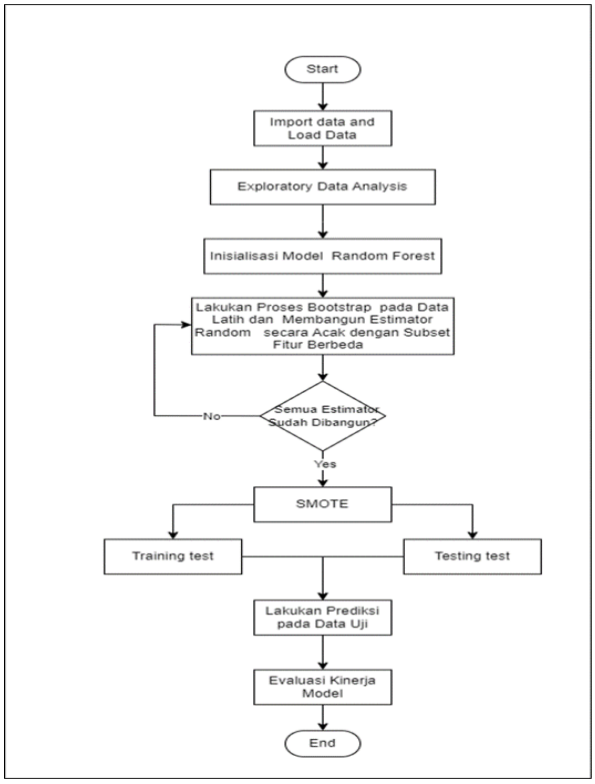
Deteksi kecurangan dalam *e-commerce* adalah penerapan metode analisis data dan pembelajaran mesin untuk mengidentifikasi, mencegah, dan mengurangi tindakan penipuan dalam transaksi *online*. Ini termasuk pembayaran palsu, pencurian identitas, penyalahgunaan kartu kredit, peninjauan palsu, dan penipuan afiliasi. Tujuannya adalah untuk menjaga integritas platform *e-commerce*, pembeli, dan penjual dengan menganalisis pola perilaku pengguna, riwayat transaksi, dan metadata.[13]

METODE PENELITIAN

Metode penelitian yang akan digunakan adalah teknik *machine learning* algoritma *Random Forest* . Data bersifat sekunder yang mana penulis melakukan pengumpulan data melalui studi literatur. Data ini kemudian diolah dan dianalisis menggunakan algoritma *Random Forest* untuk mengidentifikasi pola transaksi yang mencurigakan. Dengan cara mengumpulkan data-data yang berhubungan dengan objek penelitian, peneliti dapat lebih mudah mengidentifikasi masalah-masalah yang terkait. Alur penelitian ini dijelaskan dalam Gambar 1.

Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah dataset transaksi *e-commerce* yang mencakup fitur-fitur penting. Data ini diambil dari *website kaggle* dalam bentuk CSV. Dataset ini berisi 23,634 entri, masing-masing dengan 16 kolom yang menjelaskan transaksi *e-commerce* dan fitur yang terkait yang digunakan untuk mendeteksi penipuan. dapat dilihat pada Tabel 1 deskripsi dataset



Gambar 1. Alur Penelitian

Tabel 1. Deskripsi Dataset

No	Atribut	Deskripsi
1	*Transaction ID*	Identifikator unik untuk setiap transaksi
2	*Customer ID*	Identifikator unik untuk setiap pelanggan
3	*Transaction Amount*	Jumlah total uang yang dipertukarkan dalam transaksi
4	*Transaction Date*	Tanggal dan waktu transaksi terjadi
5	*Payment Method*	Metode yang digunakan untuk menyelesaikan transaksi
6	*Product Category*	Kategori produk yang terlibat dalam transaksi
7	*Quantity*	Jumlah produk yang terlibat dalam transaksi
8	*Customer Age*	Usia pelanggan yang melakukan transaksi
9	*Customer Location*	Lokasi geografis pelanggan
10	*Device Used*	Jenis perangkat yang digunakan untuk melakukan transaksi
11	*IP Address*	Alamat IP perangkat yang digunakan untuk transaksi
12	*Shipping Address*	Alamat pengiriman produk
14	*Is Fraudulent*	Indikator biner apakah transaksi curang (1 untuk curang, 0 untuk sa
15	*Account Age Days*	Usia akun pelanggan dalam hari pada saat transaksi
16	*Transaction Hour*	Jam dalam sehari ketika transaksi terjadi

Dengan struktur data yang komprehensif ini, penelitian dapat melakukan analisis mendalam untuk mengidentifikasi pola kecurangan yang kompleks dalam transaksi *e-commerce*, memanfaatkan kekuatan *Random Forest* dalam menangani dataset multidimensi.

Pra-pemprosesan Data

Fase ini melibatkan pengolahan dan analisis data yang dikumpulkan, termasuk penghapusan data yang tidak relevan, pengelompokan data, dan penghitung data. Data dibersihkan dan diubah dengan mengatasi nilai yang tidak ada, menghapus duplikasi, normalisasi fitur numerik, dan *encoding* fitur kategorik agar algoritma *Random Forest* dapat menanganinya. Selanjutnya pemilihan fitur, proses memilih fitur untuk mendeteksi kecurangan, ciri-ciri seperti pola transaksi, lokasi, waktu, dan lain-lain harus diidentifikasi. Memilih fitur yang paling penting dengan menggunakan metode seleksi fitur seperti analisis korelasi, uji statistik, atau metode *wrapper/filter*.

Pembagian Data

Model *Random Forest* kemudian dilatih dengan menggunakan data latih yang telah disiapkan. Disini data dibagi menjadi data pelatihan (*train*) dan data pengujian (*test*) dengan rasio, 80% untuk *train* data dan 20% untuk *test* data.

Perancangan Model

Fase ini mencakup proses pengembangan hipotesis, definisi variabel, dan perancangan algoritma yang akan digunakan untuk menganalisis data. Implementasikan algoritma *Random Forest* pada set data

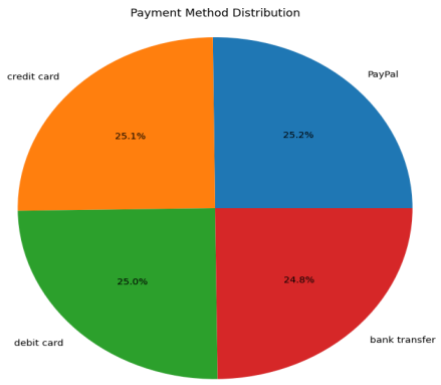
pelatihan. *Tuning hyperparameter* untuk mengoptimalkan kinerja model. Ini termasuk kriteria seperti jumlah pohon, kedalaman maksimum pohon, dan kriteria pemisahan.

Pelatihan dan Evaluasi:

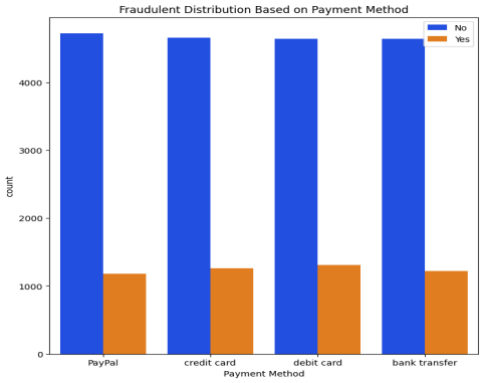
Fase ini melibatkan penggunaan data yang dikumpulkan untuk mengevaluasi kinerja model dan proses pelatihannya. Mengevaluasi kinerja model *Random Forest* menggunakan metrik yang sesuai, seperti akurasi, presisi, *recall*, skor F1. Kemudian menganalisis hasil evaluasi untuk memahami kekuatan dan kelemahan model.

HASIL DAN PEMBAHASAN

Hasil eksperimen menunjukkan bahwa dari *dataset* ini menunjukan pola transaksi *e-commerce* yang tipikal dan bervariasi. Menyoroti poin-poin penting atau temuan utama yang ditunjukkan diagram dapat dijelaskan pada Gambar 2. Pada gambar 2 dapat disimpulkan bahwa Metode pembayaran *PayPal* berwarna biru 25.2%, *credit card* berwarna oranye 25.1%, *debit card* berwarna hijau 25.0%, bank transfer berwarna merah 24.8% dari total transaksi. keempat metode pembayaran memiliki proporsi yang hampir sama dalam distribusi penggunaan. *PayPal* sedikit lebih sering digunakan, sementara *bank transfer* sedikit paling jarang digunakan. Seluruh metode pembayaran ini memiliki *persentase* yang sangat dekat satu sama lain, menunjukkan keseimbangan dalam preferensi metode pembayaran di antara pengguna *e-commerce*.

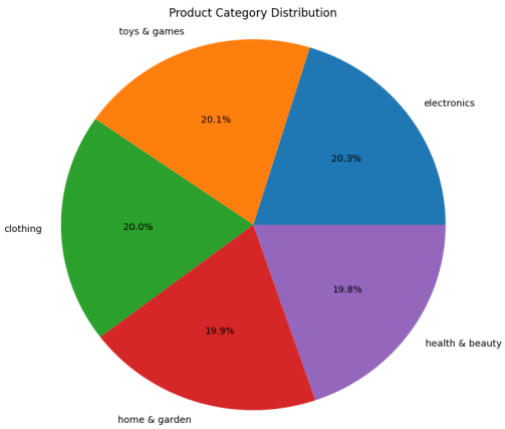


Gambar 2. *Payment Method Distribution*

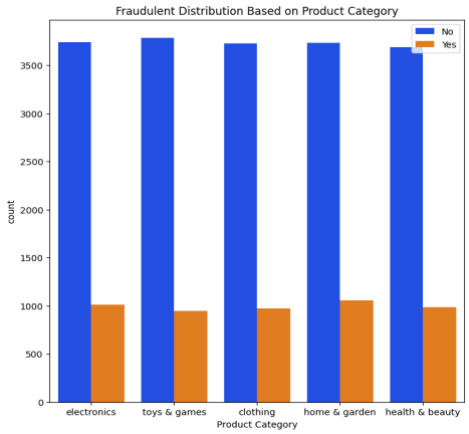


Gambar 3. Fraudulent Based on Payment Method

Pada gambar 3 ini menggambarkan distribusi transaksi curang dan tidak curang berdasarkan metode pembayaran di situs *e-commerce*. Di antara semua metode, *debit card* menunjukkan jumlah transaksi curang tertinggi, meskipun transaksi tidak curang masih jauh lebih dominan untuk semua metode pembayaran. Meskipun ada sedikit variasi, jumlah transaksi curang relatif konsisten di semua metode pembayaran, berkisar antara 1000 hingga 1300 kasus. Pada gambar 4 menunjukkan distribusi kategori produk di situs *e-commerce* memberikan wawasan penting terkait area-area yang mungkin rentan terhadap kecurangan. Dengan kategori produk yang terdistribusi merata *electronics* (20.3%), *toys & games* (20.1%), *clothing* (20.0%), *home & garden* (19.9%), dan *health & beauty* (19.8%) dengan memahami distribusi ini, model dapat dikalibrasi untuk mendeteksi anomali dalam pola pembelian di kategori dengan volume transaksi tinggi atau rendah. Dengan demikian, analisis ini akan membantu dalam meminimalkan risiko penipuan di berbagai kategori produk, meningkatkan keamanan dan kepercayaan pengguna di *platform e-commerce*.



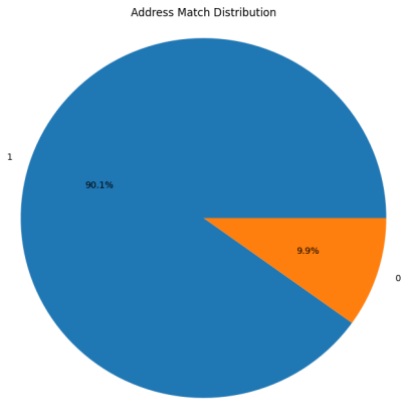
Gambar 4. *Product Category Distribution*



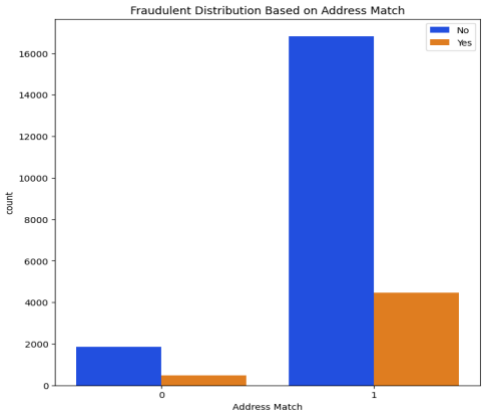
Gambar 5. ***Fraudulent Based on Product Category***

Pada gambar 5 ini menggambarkan distribusi transaksi curang dan tidak curang di *platform e-commerce* berdasarkan kategori produk. Dari lima kategori yang ditampilkan *electronics, toys & games, clothing, home & garden, dan health & beauty* terlihat bahwa kategori *home & garden* menunjukkan jumlah transaksi curang tertinggi, sementara *toys & games* memiliki jumlah terendah. Secara keseluruhan, data menunjukkan bahwa risiko penipuan ada di semua kategori produk. Meskipun kecurangan relatif jarang terjadi, perbedaan ini mengindikasikan bahwa beberapa kategori produk mungkin lebih rentan terhadap aktivitas curang, dengan jumlah transaksi curang berkisar sekitar 1000 kasus untuk setiap kategori. Namun memberikan wawasan berharga untuk strategi pencegahan penipuan yang lebih terfokus.

Pada gambar 6 menunjukkan distribusi kecocokan alamat dalam transaksi di situs *e-commerce*. Dari total transaksi, 90.1% memiliki alamat pengiriman yang cocok, sementara 9.9% memiliki ketidakcocokan alamat. Kecocokan alamat merupakan indikator penting dalam mendeteksi potensi penipuan. Ketidakcocokan alamat dapat menandakan usaha untuk menyamarkan identitas atau lokasi pengguna yang sebenarnya.

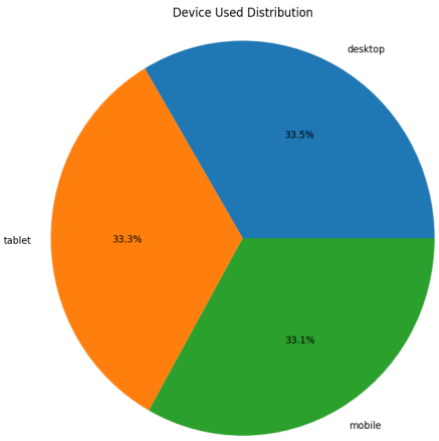


Gambar 6. ***Address match***

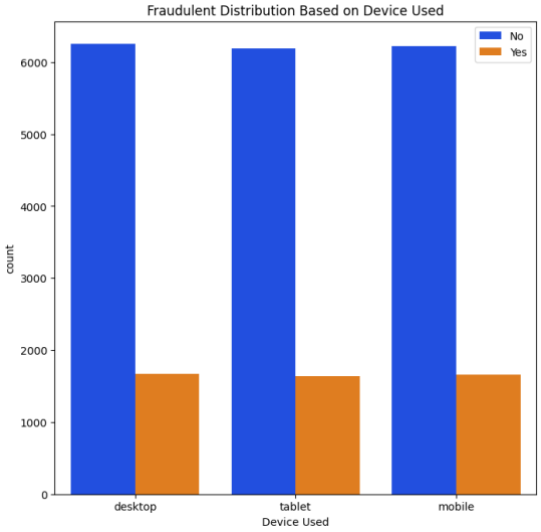


Gambar 7. ***Fraudulent based on Address Match***

Pada gambar 7 dapat diamati bahwa ketika tidak ada pencocokan alamat (0), jumlah kasus non-penipuan (biru) relatif sedikit dibandingkan dengan jumlah kasus penipuan (*oranye*). Namun, ketika terdapat pencocokan alamat (1), jumlah kasus non-penipuan (biru) sangat besar, sedangkan jumlah kasus penipuan (*oranye*) moderat. Pada gambar 8 ini menunjukkan distribusi penggunaan perangkat berbeda, desktop, tablet, dan *mobile*. Desktop 33.5%, tablet 33.3%, dan *mobile* 33.1%. Ini mengindikasikan bahwa penggunaan ketiga jenis perangkat ini hampir sama, dengan sedikit kecenderungan lebih banyak pada penggunaan desktop. memvisualisasikan proporsi masing-masing kategori dalam keseluruhan data, memudahkan perbandingan dan analisis. menunjukkan bahwa penggunaan perangkat dalam aktivitas *e-commerce* beragam dan dapat memberikan wawasan berharga dalam mengidentifikasi pola-pola potensial terkait kecurangan.



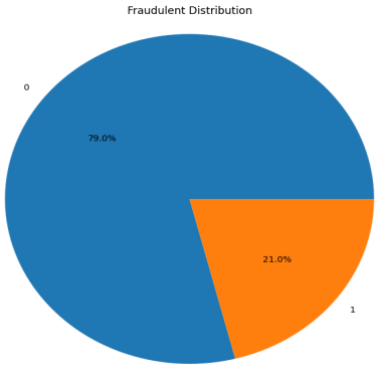
Gambar 8. ***Device Used distribution***



Gambar 9. *Fraudulent Based on Device Used*

Pada gambar 9 distribusi transaksi fraud dan non-fraud berdasarkan jenis perangkat (desktop, tablet, dan *mobile*) bahwa jumlah transaksi *non-fraud* untuk ketiga jenis perangkat mencapai sekitar 6000, sementara jumlah transaksi *fraud* sekitar 2000. Hal ini mengindikasikan bahwa meskipun proporsi antara transaksi fraud dan *non-fraud* serupa di semua perangkat, transaksi *non-fraud* jauh lebih dominan dengan rasio sekitar tiga kali lebih banyak dibandingkan transaksi *fraud*. Pada gambar 10 menggambarkan distribusi transaksi *fraud* (penipuan) dan *non-fraud*. dari total transaksi tidak ditandai sebagai *fraud* 79% dan transaksi ditandai sebagai *fraud* 21%, beberapa transaksi dianggap mencurigakan terlibat dalam kecurangan. Tabel 2 menunjukkan hasil evaluasi kinerja model dilakukan melalui penggunaan *confusion matrix*. Penilaian model ini dipertimbangkan berdasarkan tingkat akurasi, presisi, *recall*, dan f1 *score* dan *support*. Dengan akurasi 86% berarti model berhasil memprediksi dengan baik dari semua kasus yang diuji.

Untuk kelas 0, model memiliki presisi 83% dan *recall* 92%. Artinya, dari semua prediksi untuk kelas 0, 83% benar, dan model berhasil mengidentifikasi 92% dari semua kasus kelas 0 yang sebenarnya. F1-*score* untuk kelas ini adalah 0.87, yang merupakan rata-rata harmonik dari presisi dan *recall*. Kelas 1 menunjukkan presisi yang lebih tinggi (90%) namun *recall* yang lebih rendah (80%) dibandingkan kelas 0. Ini mengindikasikan bahwa model lebih akurat dalam memprediksi kelas 1, tetapi cenderung melewatkan beberapa kasus kelas 1 yang sebenarnya.



Gambar 10. *Fraudulent Distribution*

Tabel 2. Hasil Evaluasi Model

	precision	recall	f1-score	support
0	0.83	0.9	0.87	3818
1	0.90	0.8	0.85	3650
accuracy			0.86	7468
macro avg	0.86	0.8	0.86	7468
weighted av	0.86	0.8	0.86	7468

Jumlah data yang dievaluasi (*support*) cukup seimbang antara kedua kelas, dengan total 7468 sampel. Ini menunjukkan dataset yang relatif seimbang, yang baik untuk evaluasi model yang adil. Rata-rata makro dan rata-rata tertimbang keduanya menunjukkan nilai 0.86 untuk semua metrik, yang mengonfirmasi kinerja model yang konsisten di kedua kelas. Ini menandakan bahwa model memiliki performa yang baik dan seimbang dalam mengklasifikasikan kedua kelas.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian ini dapat disimpulkan bahwa algoritma *Random Forest* dapat mengidentifikasi pola kecurangan di situs *e-commerce*. Prediksi ini mencapai tingkat akurasi 86% dan menjelaskan bahwa kategori produk "*Home & Garden*" menunjukkan risiko kecurangan tertinggi, sementara "*Toys & Games*" terendah, ketidakcocokan antara alamat pengiriman dan penagihan merupakan indikator kuat potensi kecurangan dan penggunaan perangkat (desktop, tablet, *mobile*) untuk transaksi relatif seimbang dengan desktop sedikit lebih dominan. Analisis juga menunjukkan distribusi metode pembayaran yang seimbang antara *PayPal*, *credit card*, *debit crad*, dan *transfer bank*. Berdasarkan penelitian ini, disarankan untuk meningkatkan pengawasan pada kategori produk berisiko tinggi, memperketat verifikasi alamat, memantau pola transaksi berdasarkan jenis perangkat, meningkatkan keamanan pada metode pembayaran berisiko tinggi seperti *debit card*, melakukan evaluasi dan penyesuaian model

secara berkala, mengintegrasikan hasil analisis ke dalam sistem deteksi kecurangan *real-time*, dan melanjutkan penelitian untuk mengidentifikasi faktor-faktor lain yang berkontribusi pada perilaku kecurangan di *e-commerce*. Selanjutnya penelitian dapat diperluas dan diharapkan dapat secara signifikan meningkatkan keamanan dan kepercayaan dalam transaksi *e-commerce*, sambil mengurangi risiko kecurangan dan memastikan perlindungan yang lebih baik bagi pengguna *platform online.*

DAFTAR PUSTAKA

[1] R. Mustajab, "Pengguna E-Commerce RI Diproyeksi Capai 196,47 Juta pada 2023," Dataindonesia.id. [Online]. Available:https://dataindonesia.id/ekonomi-digital/detail/pengguna-ecommerce-ri-diproyeksi-capai-19647-juta-pada-2023

[2] KEMENTERIAN PERDAGANGAN RI, "Kemendag Ramal Transaksi E-Commerce di RI Tembus Rp533 Triliun," KEMENTERIAN PERDAGANGAN Repbulik Indonesia. [Online].Available:https://www.kemendag.go.id/berita/pojok-media/kemendag-ramal-transaksi-e-commerce-di-ri-tembus-rp533-triliun

[3] A. Ahdiat, "5 E-Commerce dengan Pengunjung Terbanyak Sepanjang 2023," databoks. [Online]. Available: https://databoks.katadata.co.id/datapublish/2024/01/10/5-e-commerce-dengan-pengunjung-terbanyak-sepanjang-2023

[4] C. J. Hoofnagle, B. van der Sloot, and F. Z. Borgesius, "The European Union general data protection regulation: What it is and what it means," *Inf. Commun. Technol. Law*, vol. 28, no. 1, pp. 65–98, 2019,doi:10.1080/13600834.2019.1573501.

[5] M. K. G, "Accuracy Analysis for Logistic Regression Algorithm and Random Forest Algorithm to Detect Frauds in Mobile Money Transaction," *Rev. Gestão Inovação e Tecnol.*, vol. 11, no. 4, pp. 1228–1240, 2021, doi: 10.47059/revistageintec.v11i4.2182.

[6] K. C. Laudon and C. G. Traver, "E - commerce".

[7] Rahmati. 2009, "PEMANFAATAN E-COMMERCE DALAM BISNIS DI INDONESIA."

[8] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Stat. Sci.*, vol. 17, no. 3, pp. 235–255, 2002, doi: 10.1214/ss/1042727940.

[9] H. C. Marwi and I. Oskar, "Analysis of Increasing Types of Online…. Analysis of Increasing Types of Online Fraud and Level of Public Awareness in Indonesia," vol. 4, no. November, pp. 70–84, 2023.

[10] Suci Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa

Bantaeng," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 121–127, 2022, doi: 10.35580/variansiunm31.

[11] D. T. Ananto *et al.*, "Edukasi dan Pelatihan Pengenalan Machine Learning dan Computer Vision Untuk Mengeksplorasi Potensi Visual," *Pros. Semin. Nas. Pengabdi. Masy. LPPM UMJ*, vol. 1, no. 1, pp. 1–8, 2023, [Online].Available:https://jurnal.umj.ac.id/index.php/semnaskat/article/view/19491

[12] S. J. Russell *et al.*, "Artificial Intelligence".

[13] V. Van Vlasselaer *et al.*, "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions," *Decis. Support Syst.*, vol. 75, pp. 38–48, 2015, doi: 10.1016/j.dss.2015.04.013.

[10] Purnama Ramadani Silalahi1 , Aisy Salwa Daulay2 , Tanta Sudiro Siregar3 , Aldy Ridwan4, Analisis Keamanan Transaksi *E-commerce* Dalam Mencegah Penipuan Online Jurnal Manajemen, Bisnis dan Akuntansi Vol.1, No.4 November 2022 e-ISSN: 2963-5292; p-ISSN: 2963-4989, Hal 224-235

[11] C. Tejasri*1, CH Sai Ushanth Aryan*2, D. Deekshith*3, Arrolla Chintu*4, FRAUD DETECTION IN *E-COMMERCE* USING MACHINE LEARNING e-ISSN: 2582-5208 International Research Journal of Modernization in Engineering Technology and Science Volume:04/Issue:06/June-2022 Impact Factor- 6.752 www.irjmets.com

[12] Adi Saputra1 , Suharjito2 Fraud Detection using Machine Learning in *e-commerce* (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019

[13] Neha Purohit and Dr. Rajeev G. Vishwakarma "Credit Card Fraud Detection Using Machine Learning Algorithms Using Python" Technology Webology (ISSN: 1735-188X) Volume 18, Number 6, 2021

[14] https://www.cnbcindonesia.com /tech/20230302140853-37418315/korban-penipuan-ecommerce-ri-makin-banyak-cek-data-terbaru

[15] Merih Bozbura1 , Hunkar C. Tunc2 , Miray Endican Kusak1 and C. Okan Sakar3 "Detection of *e-commerce* Anomalies using LSTM-recurrent Neural Networks" DOI: 10.5220/0007924502170224 In Proceedings of the 8th International Conference on Data Science, Technology and Applications (DATA 2019), pages 217-224 ISBN: 978-989-758-377-3

[16] Erlina Permata Sari, Deyana Annisa Febrianti, Riska Hikmah Fauziah "Fenomena Penipuan Transaksi Jual Beli Online Melalui Media Baru

Berdasarkan Kajian Space Transition Theory" DEVIANCE JURNAL KRIMINOLOGI Volume 6 Nomor 2 Desember 2022 Hal: 153-168 DOI: http://dx.doi.org/10.36080/djk.1882

[17] Paulin K. Kamuangu "A Review on Financial Fraud Detection using AI and Machine" Journal of Economics, Finance and Accounting Studies ISSN: 2709-0809 DOI: 10.32996/jefas