Prosiding Seminar Nasional Ilmu Teknik
Volume. 2 Nomor. 2, Desember 2025
E-ISSN .: 3063-5713; P-ISSN .: 3063-4709; Hal. 472-483
DOI: https://doi.org/10.61132/prosemnasproit.v2i2.152
Tersedia: https://prosiding.aritekin.or.id/index.php/PROSEMNASPROIT

Evaluasi Kinerja Model Klasifikasi Machine Learning untuk Deteksi Fraud pada Data Transaksi Tidak Seimbang

Kaslin Yulianty [1*], Dodo Zaenal Abidin [2], Joni Devitra [3]
[1-3] Magister Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa Jambi, Indonesia
Email: kaslianty@gmail.com [1], dodozaenalabidin@gmail.com [2], devitrajoni@yahoo.co.id [3]*

**Penulis Korespondensi: kaslianty@gmail.com*

Abstract. *Credit card fraud detection faces extreme class imbalance, so performance evaluation cannot rely on accuracy alone. This study constructs a comparative baseline on the public Credit Card Fraud Detection dataset with the binary label Class (0 = non-fraud, 1 = fraud). The data are processed through quality checks, duplicate removal, feature–target separation, and an 80:20 stratified train–test split. To reduce majority-class bias, SMOTE is applied only to the training set, while the test set is kept in its original distribution. Four baseline models—Logistic Regression, LinearSVC, Random Forest, and XGBoost—are compared to characterize the trade-off between detection sensitivity and false alarms on the original test set. Preprocessing is tailored to model characteristics: StandardScaler is used for linear models (Logistic Regression and LinearSVC), whereas Random Forest and XGBoost are trained without scaling. Evaluation includes a confusion matrix, fraud-class (precision/recall/F1-score), Balanced Accuracy, ROC-AUC, and ROC curves. Results show ROC-AUC around 0.96 across all models, while revealing clear trade-offs between False Positives and False Negatives; therefore, reporting minority-class metrics and FP–FN patterns is essential for representative evaluation on highly imbalanced data.*

Keywords: *Balanced Accuracy, Credit Card Transactions, Fraud Detection, ROC-AUC, SMOTE.*

Abstrak. Deteksi *fraud* pada transaksi kartu kredit menghadapi ketidakseimbangan kelas yang ekstrem, sehingga evaluasi tidak cukup mengandalkan akurasi. Penelitian ini menyusun *baseline* komparatif pada dataset publik *Credit Card Fraud Detection* dengan label *Class* (0 = *non-fraud*, 1 = *fraud*). Data diproses melalui pemeriksaan kualitas, penghapusan duplikasi, pemisahan fitur–target, dan *stratified train–test split* 80:20. Untuk mengurangi bias kelas mayoritas, SMOTE diterapkan hanya pada data latih, sedangkan data uji dipertahankan pada distribusi asli. Penelitian ini membandingkan empat model *baseline*, yaitu *Logistic Regression*, *LinearSVC*, *Random Forest*, dan *XGBoost*, untuk memetakan kompromi sensitivitas deteksi dan alarm palsu pada data uji asli. Praproses disesuaikan dengan algoritma: *StandardScaler* digunakan untuk model linear (*Logistic Regression* dan *LinearSVC*), sementara *Random Forest* dan *XGBoost* dijalankan tanpa *scaling*. Evaluasi pada data uji mencakup *confusion matrix*, metrik kelas *fraud* (*precision/recall/F1-score*), *Balanced Accuracy*, dan ROC-AUC, serta kurva ROC. Hasil menunjukkan ROC-AUC sekitar 0,96 pada seluruh model, namun terdapat *tradeoff* yang jelas antara *False Positive* dan *False Negative*, sehingga pelaporan metrik kelas minoritas dan pola FP–FN penting untuk evaluasi yang representatif pada data sangat tidak seimbang.

Kata kunci: *Balanced Accuracy*, Deteksi *Fraud*, ROC-AUC, SMOTE, Transaksi Kartu Kredit.

1. LATAR BELAKANG

Peningkatan kasus *fraud* pada perbankan, layanan keuangan, dan transaksi kartu kredit memperbesar risiko pada sistem pembayaran digital (Hafez et al., 2025). Laporan ACFE juga menunjukkan bahwa *fraud* menimbulkan kerugian finansial yang signifikan bagi berbagai organisasi (Association of Certified Fraud Examiners, 2024). Pada pembayaran daring, pola serangan yang dinamis membuat metode deteksi berbasis aturan menjadi kurang adaptif, sehingga dibutuhkan pendekatan yang mampu mengikuti perubahan pola (Ibrahim & Alfauzan,

Naskah Masuk: 30 November 2025; Revisi: 29 November 2025; Diterima: 28 Desember 2025; Terbit: 31 Desember 2025

2025). Deteksi *credit card fraud* banyak digunakan sebagai studi kasus karena volume transaksi yang besar dan kompleks, serta konsekuensi kesalahan deteksi yang dapat memengaruhi perlindungan aset pemegang kartu sekaligus meningkatkan beban operasional lembaga keuangan (Xie et al., 2021). Tantangan utamanya adalah ketidakseimbangan kelas yang ekstrem, yaitu transaksi *fraud* jauh lebih sedikit dibanding transaksi normal (Feng & Kim, 2024). Kondisi ini berpotensi membuat pemodelan dan evaluasi bias terhadap kelas mayoritas, sehingga performa pada kelas minoritas tidak tercermin memadai jika hanya mengandalkan metrik umum (Alothman et al., 2022). Oleh karena itu, evaluasi perlu menekankan metrik yang lebih representatif untuk data tidak seimbang.

Seiring berkembangnya modus *fraud*, pendekatan *machine learning* dimanfaatkan untuk memproses data transaksi berskala besar dan menyesuaikan diri terhadap pola baru (Baisholan et al., 2025). Namun, berbagai penelitian menunjukkan bahwa hasil deteksi dapat bervariasi akibat perbedaan pemilihan model, strategi penanganan *data imbalance*, dan metrik evaluasi, sehingga muncul *trade-off* yang berdampak pada biaya operasional (Ibrahim & Alfauzan, 2025), (Malek et al., 2023) (Mînăstireanu & Meșniță, 2020), (Xie et al., 2021). Berdasarkan uraian tersebut, artikel ini mengusulkan studi *baseline* komparatif pada dataset *credit card fraud detection* yang *highly imbalanced* dengan membandingkan *Logistic Regression*, *Linear SVM (LinearSVC)*, *Random Forest*, dan *XGBoost*. Kontribusi artikel adalah menyediakan *pipeline* pemrosesan dan evaluasi yang konsisten, serta menekankan metrik yang lebih sesuai untuk data tidak seimbang misalnya *precision/recall/F1-score* kelas *fraud*, *Balanced Accuracy*, dan ROC-AUC guna mendukung analisis *trade-off* kesalahan FP dan FN.

2. KAJIAN TEORITIS

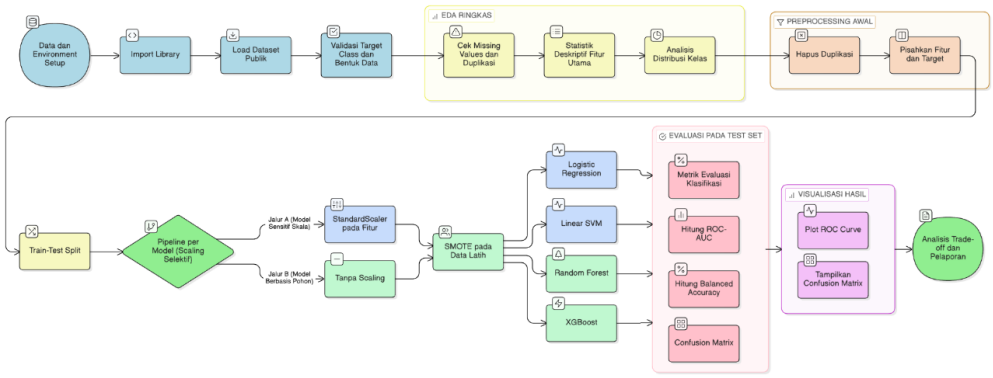
Secara umum, *fraud* merupakan tindakan penipuan yang dilakukan secara sengaja melalui misrepresentasi, penyembunyian fakta penting, atau penyalahgunaan kepercayaan, sehingga merugikan pihak lain dan memberi keuntungan tidak sah bagi pelaku (Black, 1910). Dalam konteks pelaporan keuangan, *fraud* dapat muncul sebagai *fraudulent financial statements*, yaitu manipulasi elemen keuangan seperti pendapatan, aset, penjualan, atau laba, maupun perendahan biaya, kewajiban, atau kerugian, sehingga menghasilkan informasi yang menyesatkan (Ashtiani & Raahemi, 2021). Pada ekosistem digital, bentuk *fraud* juga mencakup penipuan transaksi, penyalahgunaan identitas, dan rekayasa sosial, yang menunjukkan bahwa pola serangan dapat beragam dan berkembang.

Istilah *confidence game (con)*, *scam*, dan *scheme* digunakan untuk menggambarkan variasi mekanisme penipuan, mulai dari pembentukan kepercayaan korban hingga rencana penipuan yang terstruktur (Yunita et al., 2023), (Zhang & Dong, 2023). Dalam praktiknya, pelaku dapat menggabungkan aktivitas *online* dan *offline* sehingga serangan makin kompleks dan memerlukan deteksi yang adaptif (Lusthaus et al., 2024). Perkembangan transaksi digital mendorong kebutuhan deteksi yang lebih adaptif, sejalan dengan pemanfaatan big data, *cloud computing*, serta *Artificial Intelligence* (AI) melalui *machine learning* (ML) untuk analitik prediktif dan pemantauan transaksi yang lebih responsif (Shetty et al., 2023). Pendekatan ML mendukung analitik prediktif dan pemantauan transaksi yang lebih responsif, serta dapat diperbarui seiring bertambahnya data sehingga lebih tahan terhadap perubahan pola dibanding metode konvensional yang statis (Bello et al., 2023). Di ranah *fraud* finansial, keterbatasan pemeriksaan manual menjadi tantangan ketika volume data sangat besar; misalnya pada audit tradisional yang cenderung mahal, memakan waktu, dan tidak selalu akurat, sehingga pendekatan berbasis ML dan *data mining* banyak diteliti untuk meningkatkan efektivitas deteksi (Rabade, 2022).

Dalam konteks klasifikasi, ketidakseimbangan kelas merupakan permasalahan umum pada dataset dunia nyata dan berdampak signifikan terhadap kinerja model, terutama ketika kelas minoritas sangat kurang terwakili (Jegierski & Saganowski, 2020). Pada kondisi tersebut, model dapat tampak baik secara akurasi tetapi lemah dalam mengenali kelas minoritas, sehingga strategi penanganan data tidak seimbang menjadi krusial agar pembelajaran tidak bias terhadap kelas mayoritas. Salah satu teknik yang banyak digunakan adalah SMOTE (*Synthetic Minority Over-sampling Technique*), yaitu metode *oversampling* yang menghasilkan sampel sintetis kelas minoritas melalui interpolasi antar data minoritas yang berdekatan untuk meningkatkan representasi kelas minoritas pada data latih (Elreedy et al., 2024).

3. METODE PENELITIAN

Penelitian ini merupakan studi kuantitatif berbasis eksperimen komputasional menggunakan dataset publik *Credit Card Fraud Detection* dengan target *Class* untuk membedakan transaksi *fraud* dan *non-fraud*. Dataset memuat 284.807 transaksi kartu kredit pemegang kartu Eropa yang terjadi selama dua hari pada September 2013, dengan 492 transaksi teridentifikasi sebagai *fraud* (0,173%), sehingga bersifat *highly imbalanced* (ULB, 2021).



Gambar 1. Metode Penelitian.

Sumber: Data Olahan Peneliti (2026).

Gambar 1 merangkum alur penelitian: data dimuat dan divalidasi (pemeriksaan target *Class* dan bentuk data), dilanjutkan EDA ringkas untuk mengecek nilai hilang, duplikasi, dan distribusi kelas yang *highly imbalanced*. Setelah duplikasi dihapus, fitur (X) dan target (y) dipisahkan, kemudian data dibagi menggunakan *stratified train–test split* agar proporsi kelas minoritas tetap terjaga pada data latih dan uji. Penanganan ketidakseimbangan dilakukan dengan SMOTE yang diterapkan hanya pada data latih untuk mencegah *data leakage*, sementara data uji dipertahankan pada distribusi asli. Tahap praproses disesuaikan dengan karakter algoritma: *StandardScaler* digunakan pada *Logistic Regression* (*LogReg*) dan *Linear SVM (LinearSVC)*, sedangkan *Random Forest* (RF) dan *XGBoost* dijalankan tanpa *scaling*. Seluruh model dievaluasi pada data uji menggunakan *precision*, *recall*, dan *F1-score* untuk kelas *fraud*, serta *balanced accuracy*, ROC-AUC, dan *confusion matrix*; hasilnya divisualisasikan melalui kurva ROC untuk mendukung analisis trade-off FP dan FN.

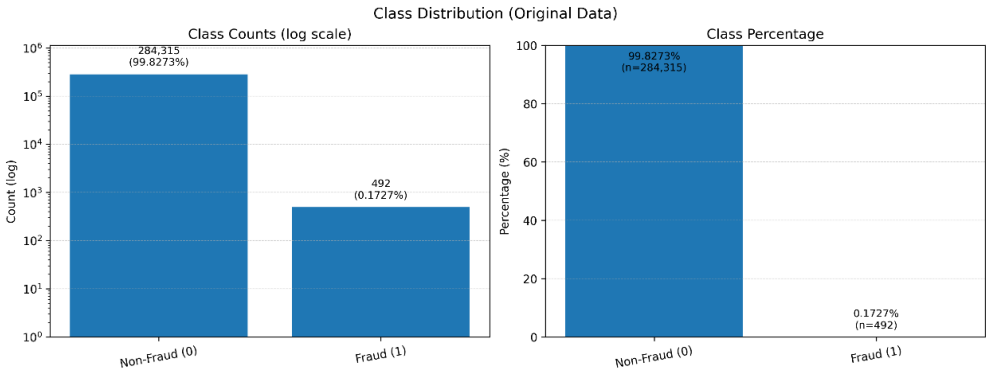
4. HASIL DAN PEMBAHASAN

Bagian ini merangkum hasil eksperimen *baseline* pada dataset *Credit Card Fraud Detection*. Proses analisis dilakukan secara berurutan mulai dari pemeriksaan data, pembentukan data latih–uji, penanganan ketidakseimbangan kelas, pelatihan model, hingga evaluasi pada data uji. Pembahasan difokuskan pada konsekuensi kesalahan *False Positive* (FP) dan *False Negative* (FN) pada konteks data yang sangat tidak seimbang.

Karakteristik Data dan Pemeriksaan Awal

Dataset yang digunakan bersifat *highly imbalanced* (492 *fraud* dari 284.807 transaksi) dengan target *Class* (0 = *non-fraud*, 1 = *fraud*), sehingga evaluasi difokuskan pada metrik kelas minoritas dan analisis FP–FN. Seluruh fitur berupa numerik dan telah dianonimkan.

Pemeriksaan awal dilakukan untuk memastikan kualitas data sebelum pemodelan, meliputi pengecekan nilai hilang, duplikasi, serta visualisasi distribusi kelas.



Gambar 2. Distribusi Kelas.

Sumber: Data Olahan Peneliti (2026).

Gambar 2 menampilkan distribusi kelas pada data awal sebelum tahapan pembersihan duplikasi, sehingga menggambarkan tingkat ketimpangan kelas yang sangat jelas antara transaksi non-*fraud* dan *fraud* pada kondisi dataset asli. Pada panel kiri (skala log), jumlah transaksi *non-fraud* jauh mendominasi (284.315; 99,8273%) dibanding transaksi *fraud* (492; 0,1727%), sehingga perbedaan jumlah tetap terlihat meskipun menggunakan skala log. Panel kanan menegaskan ketimpangan yang sama dalam bentuk persentase, di mana kelas *fraud* hampir mendekati 0% dari total data. Kondisi *highly imbalanced* ini membuat evaluasi tidak cukup mengandalkan akurasi, sehingga analisis difokuskan pada metrik kelas *fraud* serta *trade-off* kesalahan FP dan FN yang ditunjukkan melalui *confusion matrix*.

Pembentukan Data Latih–Uji dan Pencegahan *Leakage*

Setelah pembersihan duplikasi, data dipisahkan menjadi fitur (X) dan target (y). Dataset kemudian dibagi menjadi 80% data latih dan 20% data uji dengan pendekatan stratifikasi agar proporsi kelas tetap terjaga. Seluruh pemrosesan (*scaling* maupun *oversampling*) dilakukan dengan prinsip fit/transform berbasis data latih, sehingga data uji tetap menjadi acuan evaluasi yang independen.

Penanganan Ketidakseimbangan Kelas dan Alur *Preprocessing*

Ketidakseimbangan kelas ditangani menggunakan SMOTE pada data latih saja, sedangkan data uji dipertahankan pada distribusi aslinya. Pendekatan ini menjaga evaluasi tetap realistis sekaligus mencegah kebocoran informasi akibat perubahan distribusi pada data uji. *Preprocessing* dibedakan menjadi dua jalur sesuai karakter model:

1. *Logistic Regression* dan Linear *SVM/LinearSVC*: menggunakan *StandardScaler* berbasis data latih, lalu dilakukan SMOTE pada data latih yang telah dinormalisasi.

2. *Random Forest* dan *XGBoost*: tidak menggunakan *scaling*, namun SMOTE tetap hanya diterapkan pada data latih.

Model *Baseline* dan Skema Reproduksibilitas

Eksperimen membandingkan empat *baseline*: *Logistic Regression*, *LinearSVC*, *Random Forest* dan *XGBoost*. Model dan *scaler* disimpan menggunakan *joblib* sehingga eksperimen dapat dijalankan ulang secara konsisten (*TRAIN* saat belum ada file, *LOAD* saat file tersedia). Untuk kebutuhan ROC, model probabilistik menggunakan skor berbasis probabilitas kelas *fraud*, sedangkan *LinearSVC* menggunakan skor margin keputusan.

Hasil Evaluasi pada Data Uji

Evaluasi dilakukan pada data uji asli dan disajikan dalam bentuk tabel serta visualisasi. Untuk tiap model dihitung komponen TN, FP, FN, TP, diikuti metrik: *Accuracy*, *Precision/Recall/F1-score* (*Fraud)*, *Balanced Accuracy*, dan *ROC-AUC*, serta *classification report* sebagai rincian performa per kelas.

Tabel 1. Perbandingan Kinerja *Logistic Regression*, *LinearSVC, Random Forest* dan *XGBoost.*

Model	***Class***	***Precision***	***Recall***	***Fi-Score***	***Support***
Logistic Regression	*Non-Fraud*	0.9998	0.9738	0.9866	56651
	Fraud	0.0530	0.8737	0.1000	95
	Accuracy			0.9737	56746
	Macro Avg	0.5264	0.9238	0.5433	56746
	Weighted Avg	0.9982	0.9737	0.9852	56746
LinearSVC	*Non-Fraud*	0.9998	0.9803	0.9900	56651
	Fraud	0.0693	0.8737	0.1284	95
	Accuracy			0.9801	56746
	Macro Avg	0.5345	0.9270	0.5592	56746
	Weighted Avg	0.9982	0.9801	0.9885	56746
Random Forest	*Non-Fraud*	0.9996	0.9998	0.9997	56651
	Fraud	0.8795	0.7684	0.8202	95
	Accuracy			0.9994	56746
	Macro Avg	0.9396	0.8841	0.9100	56746
	Weighted Avg	0.9994	0.9994	0.9994	56746
XGBoost	*Non-Fraud*	0.9997	0.9967	0.9982	56651
	Fraud	0.2921	0.8211	0.4309	95

Model	*Class*	*Precision*	*Recall*	*Fi-Score*	*Support*
	Accuracy			0.9964	56746
	Macro Avg	0.6459	0.9089	0.7146	56746
	Weighted Avg	0.9985	0.9964	0.9972	56746

Sumber: Data Olahan Peneliti (2026).

Tabel 1 menampilkan *classification report* empat model *baseline* pada data uji untuk kelas *Non-Fraud* dan *Fraud*, meliputi *precision*, *recall*, *F1-score*, dan *support*. Karena kelas *Non-Fraud* mendominasi data uji (56,651 sampel), seluruh model menunjukkan nilai kinerja yang sangat tinggi pada kelas ini. Oleh sebab itu, pembahasan difokuskan pada kelas *Fraud* (95 sampel) untuk menilai kemampuan deteksi pada kelas minoritas. Pada kelas *Fraud*, *Logistic Regression* dan *LinearSVC* memperlihatkan *recall* yang tinggi (0.8737) namun *precision* rendah (0.0530–0.0693), yang mengindikasikan deteksi *fraud* lebih sensitif tetapi disertai peningkatan alarm palsu (FP). *Random Forest* menunjukkan *precision* yang tinggi (0.8795) dengan *recall* 0.7684, yang mencerminkan prediksi *fraud* lebih selektif (FP lebih terkendali) namun sebagian kasus *fraud* masih terlewat (FN meningkat). *XGBoost* berada di antara kedua pola tersebut dengan *recall* 0.8211 dan *precision* 0.2921.

Tabel 2. Ringkasan Metrik Global pada Data Uji (TEST).

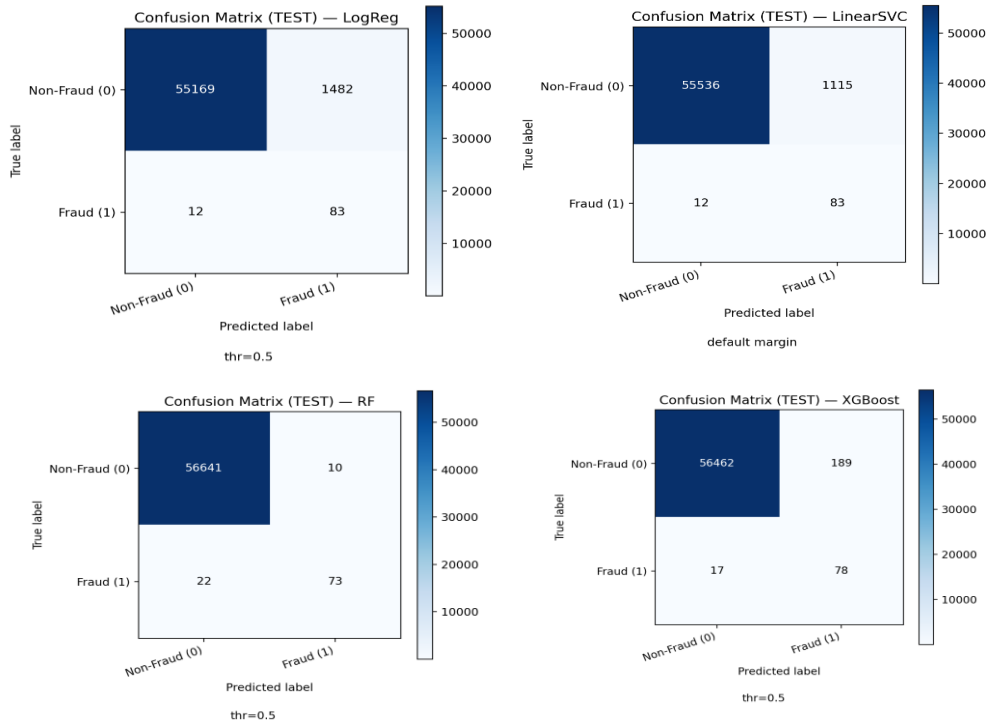
Model	Accuracy	Balanced Accuracy	ROC-AUC
Logistic Regression	0.973672	0.923762	0.962618
LinearSVC	0.980140	0.927001	0.965592
Random Forest	0.999436	0.884122	0.969009
XGBoost	0.996370	0.908858	0.958579

Sumber: Data Olahan Peneliti (2026).

Tabel 2 merangkum metrik global (*Accuracy*, *Balanced Accuracy*, dan ROC-AUC) sebagai pelengkap Tabel 1. Nilai *accuracy* tampak tinggi pada seluruh model karena dipengaruhi dominasi kelas mayoritas, sehingga interpretasi diperkuat menggunakan *balanced accuracy* yang mempertimbangkan performa kedua kelas, serta ROC-AUC yang menggambarkan kemampuan pemisahan berbasis skor prediksi. Namun, metrik global tidak secara langsung menjelaskan konsekuensi FP–FN pada satu ambang keputusan tertentu. Karena itu, pembacaan Tabel 2 tetap perlu dikaitkan dengan Tabel 1 dan *confusion matrix* untuk melihat dampak kesalahan deteksi secara praktis.

Analisis *Confusion Matrix*: *Trade-off* FP dan FN

Confusion matrix digunakan untuk meninjau konsekuensi kesalahan prediksi secara langsung. Dalam *fraud detection*, FN menunjukkan *fraud* yang terlewat, sedangkan FP menunjukkan transaksi normal yang salah ditandai *fraud*. Perbedaan strategi deteksi antar model tercermin pada perubahan FP dan FN.



Gambar 3. *Confusion matrix* (Data Uji/TEST) *Logistic Regression, LinearSVC, Random Forest* dan *XGBoost.*

Sumber: Data Olahan Peneliti (2026).

Gambar 3 memperlihatkan bagaimana masing-masing model menghasilkan kesalahan klasifikasi pada skenario data yang sangat tidak seimbang. Dalam konteks deteksi *fraud*, dua jenis kesalahan yang paling krusial adalah *False Negative* (FN), yaitu transaksi *fraud* yang diprediksi sebagai *non-fraud* (kasus *fraud* lolos), dan *False Positive* (FP), yaitu transaksi *non-fraud* yang diprediksi sebagai *fraud* (alarm palsu). Karena jumlah transaksi *non-fraud* jauh lebih besar dibanding *fraud*, nilai *True Negative* (TN) umumnya mendominasi, sehingga interpretasi performa model lebih informatif jika meninjau FN dan FP beserta konsekuensinya.

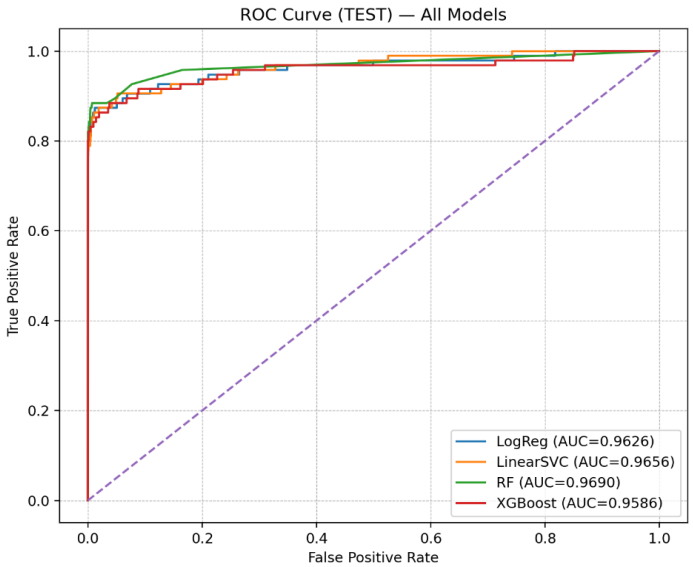
Berdasarkan Gambar 3, model linear (*Logistic Regression* dan *LinearSVC*) menunjukkan pola deteksi dengan *True Positive* (TP) relatif lebih banyak, namun disertai FP yang juga lebih besar. Pola ini sejalan dengan kecenderungan model yang lebih sensitif terhadap kelas minoritas, sehingga lebih banyak transaksi ditandai sebagai *fraud*. Sebaliknya, *Random Forest* memperlihatkan FP yang sangat kecil, tetapi masih terdapat FN yang menandakan sebagian

fraud belum terdeteksi. *XGBoost* berada di antara kedua pola tersebut, dengan FP dan FN yang tidak seekstrem dua pendekatan sebelumnya. Perbedaan ini menegaskan adanya *trade-off* antara FP dan FN yang perlu dipertimbangkan sesuai kebutuhan sistem (misalnya, meminimalkan *fraud* lolos versus mengurangi alarm palsu).

Selain itu, mekanisme pengambilan keputusan memengaruhi pembacaan hasil. *Logistic Regression*, *Random Forest*, dan *XGBoost* menggunakan probabilitas prediksi dengan ambang *threshold* thr = 0.5, sedangkan *LinearSVC* menggunakan *default margin* berbasis fungsi keputusan karena tidak menyediakan probabilitas secara bawaan. Artinya, "titik potong" keputusan pada *LinearSVC* tidak dinyatakan sebagai probabilitas 0.5, melainkan mengikuti posisi skor terhadap margin keputusan.

ROC *Curve* dan ROC-AUC

Selain evaluasi pada satu ambang keputusan tertentu, kemampuan pemisahan kelas dianalisis menggunakan kurva *Receiver Operating Characteristic* (ROC). Kurva ROC menunjukkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada berbagai *threshold* skor prediksi. Nilai ROC-AUC digunakan sebagai ringkasan numerik untuk menggambarkan kemampuan model membedakan *Fraud* dan *Non-Fraud* berdasarkan skor, tanpa bergantung pada satu *threshold*.



Gambar 4. Kurva ROC (TEST) *Logistic Regression*, *LinearSVC*, *Random Forest* dan *XGBoost*.

Sumber: Data Olahan Peneliti (2026).

Gambar 4 menampilkan kurva ROC gabungan untuk LogReg, LinearSVC, RF, dan *XGBoost* pada data uji. Garis putus-putus diagonal merepresentasikan performa acak (*random classifier*), sehingga kurva yang berada lebih jauh di atas garis tersebut mengindikasikan

pemisahan kelas yang lebih baik. Berdasarkan gambar 4, nilai *ROC-AUC* masing-masing model berada pada kisaran tinggi (LogReg = 0.9626; LinearSVC = 0.9656; RF = 0.9690; XGBoost = 0.9586), yang menunjukkan bahwa keempat model memiliki kemampuan pemisahan skor yang kuat pada data uji.

Meskipun *ROC-AUC* membantu melihat kemampuan pemisahan skor secara keseluruhan, interpretasi hasil tetap perlu dikaitkan dengan *confusion matrix* karena keputusan operasional biasanya menggunakan *threshold* tertentu (misalnya 0.5 untuk model probabilistik atau *default margin* untuk *LinearSVC*). Dengan demikian, kurva ROC berperan sebagai pelengkap untuk membaca kecenderungan model pada berbagai ambang, sementara *trade-off* FP dan FN diamati langsung melalui *confusion matrix*.

5. KESIMPULAN DAN SARAN

Penelitian ini menyusun *baseline* komparatif empat model klasifikasi pada dataset transaksi kartu kredit yang sangat tidak seimbang dengan *pipeline* praproses yang konsisten, termasuk pencegahan *data leakage* melalui pemisahan data uji sebelum transformasi serta penerapan SMOTE hanya pada data latih. Evaluasi pada data uji asli menunjukkan bahwa akurasi dapat tampak tinggi pada data *imbalanced*, sehingga interpretasi kinerja perlu menekankan metrik kelas *fraud* (*precision*/*recall*/*F1-score*), *balanced accuracy*, ROC-AUC, dan pola FP–FN pada *confusion matrix*. Untuk penelitian selanjutnya, dapat mengevaluasi pengaruh variasi ambang keputusan (*threshold*) atau margin, membandingkan strategi penanganan ketidakseimbangan lain (*oversampling*/*undersampling*) tanpa mengubah distribusi data uji, serta menambah validasi dan pelaporan keterbatasan dataset agar hasil lebih kuat dan mudah direplikasi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak yang telah memberikan dukungan selama pelaksanaan penelitian ini. Penulis juga menyampaikan apresiasi atas masukan dan saran yang diberikan selama proses penyusunan naskah ini.

DAFTAR REFERENSI

Alothman, R., AliTalib, H., & Mohammed, M. S. (2022). Fraud Detection Under the Unbalanced Class Based on Gradient Boosting. *Eastern-European Journal of Enterprise Technologies, 116*(2). https://doi.org/10.15587/1729-4061.2022.254922

Ashtiani, M. N., & Raahemi, B. (2021). Intelligent fraud detection in financial statements using machine learning and data mining: a systematic literature review. *IEEE Access, 10*, 72504-72525. https://doi.org/10.1109/ACCESS.2021.3096799

Association of Certified Fraud Examiners. (2024). *ACFE - Occupational Fraud 2024: A Report to the Nations.* https://legacy.acfe.com/report-to-the-nations/2024/

Baisholan, N., Dietz, J. E., Gnatyuk, S., Turdalyuly, M., Matson, E. T., & Baisholanova, K. (2025). A Systematic review of machine learning in credit card fraud detection under original class imbalance. *Computers, 14*(10), 437. https://doi.org/https://doi.org/10.3390/computers14100437 https://doi.org/10.3390/computers14100437

Bello, O. A., Folorunso, A., Ejiofor, O. E., Budale, F. Z., Adebayo, K., & Babatunde, O. A. (2023). Machine learning approaches for enhancing fraud prevention in financial transactions. *International Journal of Management Technology, 10*(1), 85-108. https://doi.org/https://doi.org/10.37745/ijmt.2013/vo110n185109

Black, H. C. (1910). *Law dictionary*. West Publishing Company St. Paul, Minn.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research, 16*, 321-357. https://doi.org/10.1613/jair.953

Elreedy, D., Atiya, A. F., & Kamalov, F. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning, 113*(7), 4903-4923. https://doi.org/10.1007/s10994-022-06296-4

Feng, X., & Kim, S.-K. (2024). Novel machine learning based credit card fraud detection systems. *Mathematics, 12*(12), 1869. https://doi.org/https://doi.org/10.3390/math12121869 https://doi.org/10.3390/math12121869

Gotelaere, S., & Paoli, L. (2025). Prevention and Control of Financial Fraud: a Scoping Review. *European Journal on Criminal Policy and Research, 31*(1), 1-21. https://doi.org/10.1007/s10610-022-09532-8

Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. *Journal of Big Data, 12*(1), 6. https://doi.org/https://doi.org/10.1186/s40537-024-01048-8

Ibrahim, M. M., & Alfauzan, S. (2025). Analisis Kinerja Model Machine Learning untuk Mendeteksi Transaksi Fraud pada Sistem Pembayaran Online. *JURNAL ILMIAH NUSANTARA, 2*(3), 35-49. https://doi.org/DOI: https://doi.org/10.61722/jinu.v2i3.4276

Jegierski, H., & Saganowski, S. (2020). An "outside the box" solution for imbalanced data classification. *IEEE Access, 8*, 125191-125209. https://doi.org/10.1109/ACCESS.2020.3007801

Lusthaus, J., Kleemans, E., Leukfeldt, R., Levi, M., & Holt, T. (2024). Cybercriminal networks in the UK and Beyond: Network structure, criminal cooperation and external interactions. *Trends in Organized Crime, 27*(3), 364-387. https://doi.org/10.1007/s12117-022-09476-9

Malek, N. H. A., Yaacob, W. F. W., Wah, Y. B., Nasir, S. A. M., Shaadan, N., & Indratno, S. W. (2023). Comparison of ensemble hybrid sampling with bagging and boosting machine learning approach for imbalanced data. *Indones. J. Elec. Eng. Comput. Sci, 29*(1), 598. https://doi.org/10.11591/ijeecs.v29.i1.pp598-608

Mînăstireanu, E.-A., & Meșniță, G. (2020). Methods of handling unbalanced datasets in credit card fraud detection. *Brain. Broad Research in Artificial Intelligence and Neuroscience, 11*(1), 131-143. https://doi.org/https://doi.org/10.18662/brain/11.1/19

Rabade, S. U. (2022). Use of machine learning in financial fraud detection: A review. *International Journal of Advanced Research in Science, Communication and Technology, 2*(3), 38-44. https://doi.org/10.48175/IJARSCT-7595

Shetty, V. R., R, P., & Malghan, R. L. (2023). Safeguarding against Cyber Threats: Machine Learning-Based Approaches for Real-Time Fraud Detection and Prevention. *Engineering Proceedings, 59*(1). https://doi.org/10.3390/engproc2023059111

ULB, M. L. G.-. (2021). Credit Card Fraud Detection. *Kaggle.* https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/data

Xie, Y., Li, A., Gao, L., & Liu, Z. (2021). A heterogeneous ensemble learning model based on data distribution for credit card fraud detection. *Wireless Communications and Mobile Computing, 2021*(1), 2531210. https://doi.org/https://doi.org/10.1155/2021/2531210 https://doi.org/10.1155/2021/2531210

Yunita, A., Wardhani, R. S., Levany, Y., Rahmadoni, F., Fibrianto, A., & Martoyo, A. (2023). *Manajemen Risiko Fraud.* Tohar Media.

Zhang, Y., & Dong, H. (2023). Criminal law regulation of cyber fraud crimes—from the perspective of citizens' personal information protection in the era of edge computing. *Journal of Cloud Computing, 12*(1), 64. https://doi.org/10.1186/s13677-023-00437-3