ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 14, Nomor 2, Tahun 2025, Halaman 565 - 576
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

EVALUASI MODEL KLASIFIKASI DALAM DETEKSI PENIPUAN TRANSAKSI: STUDI KASUS PADA DATA TIDAK SEIMBANG

Jefita Resti Sari[1*], Kusman Sadik[2], Agus M.Soleh[3], Cici Suhaeni[4]
[1,2,3,4] Program Studi Statistika dan Sains Data, Sekolah Sains Data, Matematika dan Informatika, Institut Pertanian Bogor
***e-mail**: jefitarestisari@apps.ipb.ac.id

DOI: 10.14710/j.gauss.14.2.565-576

Article Info:
Received: 2025-06-28
Accepted: 2025-12-05
Available Online: 2025-12-12

Keywords:
Anomly; Credit Card; SMOTE; Random Forest

Abstract: The rise in digital transactions increases the risk of credit card fraud, highlighting the need for a smart and accurate detection system. This study aims to develop a classification model that can effectively detect fraudulent transactions, despite data imbalance challenges. Data processing involves the use of the SMOTE technique to improve the representation of minority classes and Optuna for hyperparameter tuning. Three machine learning models are applied: Logistic Regression, Random Forest, and XGBoost. Model performance is evaluated using precision, recall, f1-score, and ROC-AUC. The results show that Random Forest achieves the best performance, with a precision of 0.91, recall of 0.74, and f1-score of 0.82. Logistic Regression achieves high recall but very low precision, while XGBoost produces a competitive AUC but a lower f1-score than Random Forest. This research highlights the importance of algorithm selection, data balancing with SMOTE, and parameter tuning to build an effective and adaptive fraud detection system for imbalanced data.

1. PENDAHULUAN

Kartu kredit merupakan salah satu jasa yang diberikan oleh suatu bank kepada pelanggannya. Kartu kredit sangat membantu dalam melakukan transaksi sehingga memudahkan pelanggan dalam bertransaksi tanpa memiliki uang tunai dan bisa digunakan dalam transaksi online (Alamri and Ykhlef 2024). Penipuan transaksi kartu kredit juga semakin rentan terjadi dilingkungan masyarakat dan sudah menjadi tantangan dalam industri keuangan. Penipuan transaksi anomali sudah menjadi sumber kerugian utama finansial dan sudah sering sekali terjadi (Syeeda and Mohanty 2024). Bank dan lembaga keungan lainnya sangat bekerja keras dalam menurunkan tingkat aktivitas transaksi anomali melalui pencegahan. Bank melakukan pencegahan transaksi anomali dengan berbagai cara diantaranya pemantauan skor risiko terjadinya transaksi anomali serta aturan biometrik fisik secara *real-time* (Mbakwe and Adewale 2022). Sehingga sistem perbankan sangat memerlukan metode mendeteksi transaksi yang akurat dan efisien untuk meminimalisir kerugian yang terjadi.

Proses implementasi anomali menjadi lebih sulit dideteksi bahkan dengan teknik yang terus berkembang, serta risiko anomali yang semakin merambah ke tahap aplikasi bisnis. Hal ini mengakibatkan meningkatnya jumlah aset yang bermasalah dan memberikan dampak terhadap stabilitas pasar ekonomi. Oleh karena itu, pencegahan dan deteksi anomali kartu kredit mendapat perhatian besar (Jiang et al. 2023). Tantangan dari mendeteksi transaksi anomali salah satunya yaitu kelas ketidakseimbangan (*class imbalance*) yang ekstrem pada data. Seperti dalam penelitian (F. Carcillo et al. 2019), dalam dataset transaksi kartu kredit umumnya, transaksi fraud hanya berjumlah kurang dari 1% dari total data transaksi kartu kredit. Ketidakseimbangan dalam data dapat menyebabkan model cenderung memprioritaskan kelas mayoritas (transaksi normal). Selain itu, kompleksitas pola transaksi

dan tingginya dimensi data juga merupakan tantangan dalam membangun model deteksi yang akurat dan efisien.

Mengatasi data yang tidak seimbang dalam penelitian ini, dapat dilakukan dengan menggunakan pendekatan teknik *Synthetic Minority Oversampling Technique* (SMOTE) yang dapat menyeimbangkan dataset dan menghilangkan data yang tidak jelas dan untuk meningkatkan representasi kelas minoritas tanpa menghasilkan duplikasi data yang berlebihan (Chawla, Bowyer, and Hall 2002). SMOTE terbukti mampu meningkatkan distribusi data secara lebih proporsional tanpa menambah noise secara signifikan. Dalam konteks optimasi model, pemilihan *hyperparameter* yang tepat sangat memengaruhi performa pemodelan. Optuna merupakan salah satu *framework* optimasi *hyperparameter* berbasis *Bayesian* yang dirancang untuk melakukan eksplorasi ruang parameter secara efisien (Akiba et al. 2019). Mengintegrasikan SMOTE dan Optuna, proses pelatihan model dapat dioptimalkan baik dari sisi representasi data maupun konfigurasi model, sehingga mampu mendeteksi kasus penipuan secara lebih presisi dan adaptif.

Sejumlah penelitian sebelumnya telah mengimplementasikan metode-metode yang sama. Pada penelitian (Fabrizio Carcillo et al. 2021) menunjukkan bahwa kombinasi metode *unsupervised learning* (seperti *clustering* atau *anomaly detection*) dan *supervised learning* (seperti *Random Forest, Logistic Regression*) dapat menghasilkan sistem deteksi penipuan yang tangguh dalam lingkungan data transaksi yang tidak seimbang. Penelitian (Correa Bahnsen et al. 2016) juga menekankan pentingnya teknik rekayasa fitur dan pemilihan model yang adaptif untuk meningkatkan deteksi anomali. Hasil-hasil ini memperkuat bahwa pemilihan algoritma, balancing, dan tuning merupakan elemen yang saling mendukung dalam membangun model deteksi yang efektif.

Berdasarkan penjelasan latar belakang sebelumnya, penelitian ini bertujuan untuk membangun dan membandingkan performa beberapa algoritma klasifikasi seperti *Logistic Regression, Random Forest*, dan XGBoost, dengan penerapan teknik SMOTE untuk penyeimbangan kelas serta Optuna untuk optimasi *hyperparameter* yang digunakan pada pemodelan XGBoost. Evaluasi dilakukan menggunakan metrik klasifikasi seperti *precision, recall, f1-score*, dan ROC-AUC, dengan harapan dapat mengidentifikasi model terbaik dalam mendeteksi transaksi penipuan kartu kredit secara akurat.

2. TINJAUAN PUSTAKA

Permasalahan ketidakseimbangan kelas sering terjadi dalam kasus deteksi penipuan, di mana jumlah transaksi normal jauh lebih banyak dibandingkan transaksi anomali. Ketidakseimbangan ini menyebabkan model cenderung bias terhadap kelas mayoritas. Salah satu solusi yang umum digunakan adalah teknik *Synthetic Minority Oversampling Technique* (SMOTE) yang dikembangkan oleh (Chawla, Bowyer, and Hall 2002). SMOTE bekerja dengan menghasilkan data sintetik untuk kelas minoritas melalui interpolasi linier antara sampel minoritas dan tetangganya. Teknik ini efektif dalam meningkatkan representasi kelas minoritas dan mengurangi risiko overfitting. Rumus matematis untuk pembentukan data sintetik dalam SMOTE seperti pada persamaan (1) berikut:

$$x_{new} = x_i + \lambda(x_{NN} - x_i) \tag{1}$$

dimana x_{new} sebagai titik data sintesis, x_i sebagai contoh kelas mayoritas, λ sebagai angka acak antara 0 dan 1, dan x_{NN} sebagai salah satu tetangga. Teknik ini terbukti mampu meningkatkan generalisasi model tanpa menimbulkan risiko *overfitting*, serta memberikan representasi data minoritas yang lebih baik.

Pemodelan dalam konteks pembelajaran mesin merupakan proses membangun fungsi atau struktur matematis yang merepresentasikan hubungan antara variabel bebas (fitur) dan variabel target. Tujuan utama dari pemodelan adalah untuk mengenali pola dalam data historis sehingga model dapat melakukan prediksi atau klasifikasi secara tepat pada data baru. Dalam klasifikasi, pemodelan membantu dalam mengelompokkan data ke dalam kategori yang telah ditentukan berdasarkan karakteristik input yang diamati. Keberhasilan pemodelan dipengaruhi oleh kualitas data, pemilihan fitur, dan jenis algoritma yang digunakan dalam proses pelatihan serta evaluasi model (Kotsiantis and Kanellopoulos 2006; Witten et al. 2017).

1. Random Forest

   Random Forest merupakan metode yang memiliki peran penting dalam menangani data yang kompleks. *Random forest* juga dapat menangani *overfitting* pada data. *Random forest* berfungsi untuk menggabungkan beberapa pohon keputusan sehingga dapat dimanfaatkan dalam melihat pola penipuan atau anomali pada transaksi kartu kredit (Science et al. 2024). Metode ini membentuk sejumlah besar decision tree dari subset data yang di-bootstrapped dan menggabungkan prediksinya melalui voting atau rata-rata.

   Rumus dasar prediksi *Random Forest* terdapat pada persamaan (2) berikut:

$$\hat{y} = majority\_vote(h_1(x), h_2(x), \dots, h_k(x)) \tag{2}$$

   Pada persamaan (2), $h_i(x)$ merepserentasikan fungsi prediksi dari pohon keputusan ke-I, dimana setiap pohon dibangun dari subset data dan subset fitur yang berbeda. Masing-masing $h_i(x)$ menghasilkan label prediksi untuk sampel x, kemudian model *Random Forest* menggabungkan seluruh prediksi tersebut melalui mekanisme *majority vote* sehingga kelas dengan suara terbanyak menjadi prediksi akhir. *Random Forest* unggul dalam kestabilan prediksi, tahan terhadap *overfitting*, serta efektif pada dataset berskala besar dengan fitur yang kompleks.

2. *Extreme Gradient Boosting* (XGBoosst)

   XGBoost merupakan salah satu algoritma boosting paling populer yang dikembangkan oleh Chen dan Guestrin (2016). XGBoost membangun model secara iteratif, di mana setiap model baru bertujuan memperbaiki kesalahan prediksi dari model sebelumnya. Fungsi *loss* yang dioptimasi dalam *boosting* seperti pada persamaan (3) berikut:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{3}$$

   Pada persamaan (3), y_i merupakan nilai target sebenarnya, sedangkan $\hat{y}_i^{(t-1)}$ adalah prediksi kumulatif model sebelum panambahan pohon ke-*t*. Fungsi $f_t(x_i)$ menggambarkan kontribusi pohon baru pada iterasi ke-t dalam memperbaiki kesalahan prediksi sebelumnya. Selain itu, fungsi f_t adalah model pohon pada iterasi ke-t, Ω adalah fungsi regularisasi untuk mengendalikan kompleksitas model. Kelebihan XGBoost terletak pada kecepatan, efisiensi memori, dan fleksibilitas terhadap data tidak terstruktur. XGBoost dalam pemodelannya menggunakan *Optuna Hyperparameter* dalam mengatasi atau mempersingkat waktu *tuning*.

   Optuna merupakan kerangka kerja otomatisasi pencarian *hyperparameter* berbasis pendekatan *Bayesian. Framework* ini menggunakan *Tree-structured Parzen Estimator*

(TPE) untuk mengevaluasi ruang pencarian parameter secara efisien, sekaligus menyeimbangkan eksplorasi dan eksploitasi.

Proses utama dalam Optuna terdiri dari tiga tahap:

1. Definisi ruang pencarian parameter.
2. Evaluasi nilai objektif dari parameter yang diuji.
3. Seleksi parameter terbaik berdasarkan nilai tertinggi (misalnya ROC-AUC atau F1-score).

Menurut Akiba et al. (2019), Optuna dapat mengurangi waktu *tuning* secara signifikan jika dibandingkan dengan *Grid Search* tradisional, dan dapat diintegrasikan dengan berbagai algoritma pembelajaran mesin.

3. *Logistic Regression*

Logistic Regression merupakan metode statistik klasik yang digunakan untuk memodelkan probabilitas keanggotaan suatu kelas berdasarkan nilai variabel bebas. Regresi logistik digunakan untuk memodelkan probabilitas dari suatu kejadian biner (0 atau 1), seperti halnya pada deteksi transaksi anomali atau *fraud*.

Dinyatakan melalui fungsi logit (Purwa 2019):

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}} \tag{4}$$

Model regresi logistik terbukti memberikan performa yang cukup stabil pada data yang tidak seimbang, yakni pada penelitian (Ronny Susetyoko, Wiratmoko Yuwono, and Elly Purwantini 2022) menyimpulkan bahwa regresi logistik efektif dalam menangani data yang memiliki ketidakseimbangan kelas dengan parameter sederhana dan interpretasi yang jelas. Namun, model ini mudah diinterpretasikan dan cepat dilatih, namun memiliki keterbatasan dalam menangani relasi non-linier antar fitur (Hosmer, Lemeshow, and Sturdivant 2013).

Evaluasi model klasifikasi diperlukan untuk menilai seberapa baik model dapat mengklasifikasikan kelas penipuan dan non-penipuan. Performa dari model dapat dilihat dengan menggunakan matriks sebagai berikut.

	Predicted Negatif	Predicted Positive
Actual Negative	TN	FP
Actual Positive	FN	TP

Gambar 1. *Confusion Matrix*

Beberapa metrik evaluasi penting meliputi:

- *Precision*: proporsi prediksi positif yang benar.

$$precision = \frac{TP}{TP+FP} \tag{5}$$

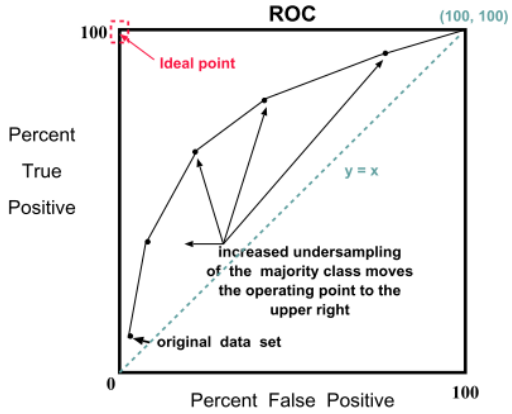
- *Recall*: proporsi kasus aktual positif yang berhasil diprediksi dengan benar.

$$recall = \frac{TP}{TP + FN} \quad (6)$$

- F1-*score*: rata-rata harmonis dari *precision* dan *recall*, digunakan saat terjadi *trade-off*.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

- ROC-AUC (*Receiver Operating Characteristic – Area Under Curve*)



Gambar 2. *Confusion Matrix* (Chawla, Bowyer, and Hall 2002)

ROC-AUC mengukur kemampuan model dalam membedakan antara kelas positif dan negatif. Nilai AUC mendekati 1 menunjukkan performa model yang baik secara keseluruhan.

3. METODE PENELITIAN

Dataset yang digunakan dalam penelitian ini merupakan data transaksi anomali kartu kredit yang tersedia di *Kaggle* : https://www.kaggle.com/code/janiobachmann/credit-fraud-dealing-with-imbalanced-datasets/ kumpulan data yang berisi transaksi yang dilakukan dengan kartu kredit pada bulan September 2013 oleh pemegang kartu Eropa. Data dalam penelitian ini memiliki beberapa variabel yang digunakan diantaranya sebagai berikut.

Tabel 1. Deskripsi Variabel Data Transaksi Anomali Kartu Kredit

Variabel	Keterangan	Tipe Data
Time	Waktu sejak transaksi pertama	Numerik
Amount	Jumlah nominal transaksi	Numerik
V1-V28	Fitur Anonim dari Hasil transformasi PCA	Numerik
Class	Pelabelan Transaksi	Kategorik 0 : Normal 1 : Anomali

- Jumlah Data: Total transaksi yang dilakukan dalam setiap transaksi dari 0 sampai dengan sebanyak 284.807 transaksi. Sebanyak 492 transaksi anomali atau 0.172% dari total transaksi.
- Variabel yang digunakan:
  1. *Time*: Waktu transaksi yang dihitung dari 0 sampai dengan 173.000 detik dalam satu hari penuh.

2. *Amount*: Fitur yang menunjukkan nilai atau nominal transaksi yang dilakukan dengan *range* yaiitu 0 sampai dengan 25.700 dollar pertransaksinya.
3. Fitur Anonim (V1-V28) : Fitur anonim sebanyak 28 fitur yang merupakan hasil dari pengurangan dimensi PCA untuk melindungi identitas pengguna dan fitur sensitif.

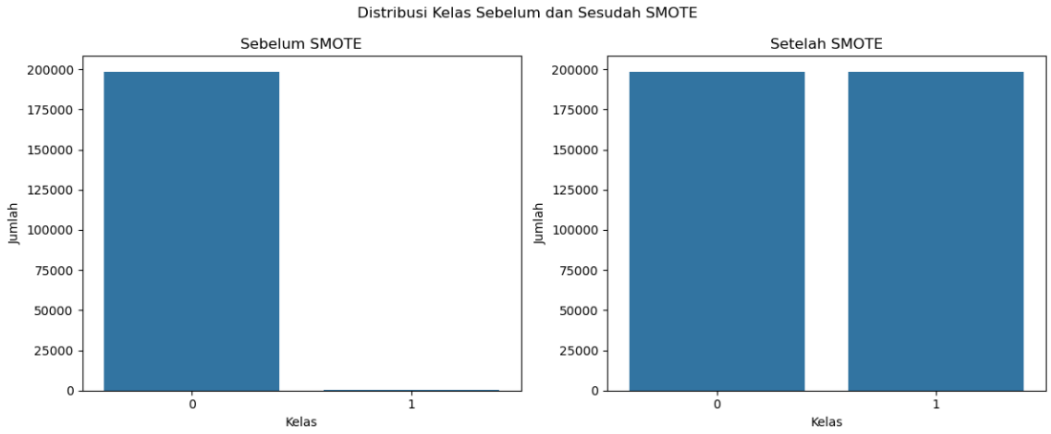
- Label (*Class*) :
  - 0 : Transaksi Normal
  - 1 : Transaksi Anomali atau *Fraud*

Prosedur analisis yang dilakukan dalam penelitian ini meliputi sebagai berikut.

1. Melakukan peng*input*an dataset yang berasal dari sumber terbuka yaitu *Kaggle* yang terdiri dari beberapa variabel yaitu *time, amount, V1-V28,* dan *class*.
2. Tahapan *Pre-processing,* data transaksi kartu kredit dieksplorasi untuk melihat statistik awal dari data, dilakukan *cleaning data*, normalisasi data, dan dilakukan proses *splitting data* yaitu 70:30.
3. Dilakukan teknik SMOTE (*Oversampling*) yang diterapkan pada data pelatihan untuk menangani ketidakseimbangan kelas, dengan cara mensintesis data minoritas (*fraud*).
4. Dilakukan pemodelan untuk melihat prediksi penipuan kartu kredit dengan menggunakan tiga model klasifikasi *machine learning*, dianataranyaL
   - *Random Forest*
     Dilatih langsung menggunakan parameter *default* tanpa proses *tuning hyperparameter*.
   - XGBoost
     Dilatih dengan data hasil SMOTE dan dioptimasi menggunakan Optuna untuk mencari kombinasi hyperparameter terbaik seperti max_depth, learning_rate, n_estimators, dan scale_pos_weight.
   - *Logistic Regression*
     Dilatih tanpa *tuning*, digunakan sebagai model dasar (*baseline*) yang sederhana dan cepat dalam pelatihan.
5. Evaluasi Model
   Setelah dilakukan pemodelan, konerja model akan diukur dengan menggunakan *confusion matrix,* metrik ROC-AUC, serta *precision, recall,* dan *F1-score* untuk menilai performa deteksi penipuan secara menyeluruh.

4. HASIL DAN PEMBAHASAN

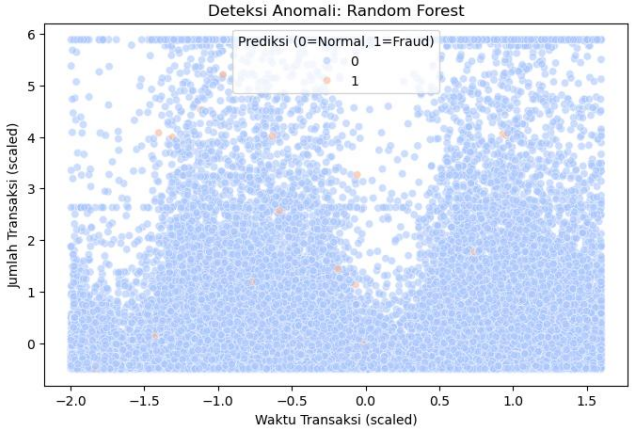
Pada tahapan *pre-processing* data penelitian dibagi menjadi data *training* dan data *testing* kemudian dilakukan tahapan dalam mengatasi ketidakseimbangan kelas dengan menggunakan SMOTE. Hal ini dapat membantu agar model dapat belajar pola penipuan yang lebih representatif dan menghindari bias ke kelas mayoritas.



Gambar 3. Distribusi kelas sebelum dan sesudah SMOTE

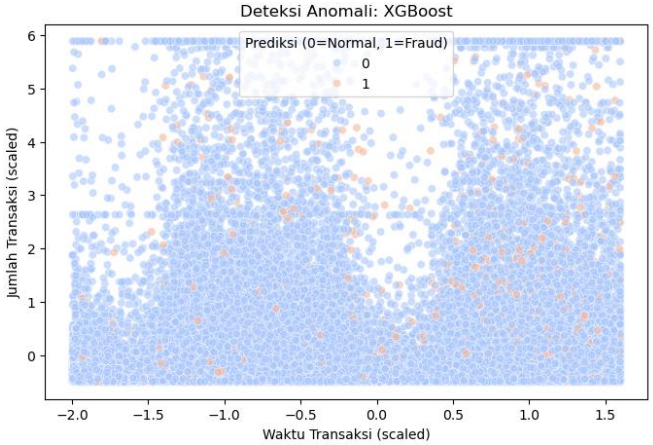
Berdasarkan Gambar 3 dapat dijelaskan bahwa penerapan metode SMOTE berhasil mengatasi ketidakseimbangan kelas dalam data pelatihan. Sebelum dilakukan SMOTE terdapat 198.277 transaksi mormal dan hanya 331 transaksi penipuan (*fraud*) yang berarti data penipuan hanya sekitar 0.17% dari keseluruhan data. Setelah dilakukan SMOTE yang diterapkan pada data pelatihan, yang menghasilkan data sintesis tambahan pada kelas minoritas (*fraud*) hingga jumlahnya sama dengan kelas mayoritas, yakni 198.277 untuk masing-masing kelas. Hasil ini menjadikan rasio distribusi kelas menjadi 50:50, yang diharapkan mampu mengurangi bias model terhadap kelas mayoritas serta meningkatkan sensitivitas model terhadap deteksi penipuan. Proses penyeimbangan data hanya diterapkan pada data pelatihan (*training set*), sedangkan data pengujian (*testing set*) tetap mempertahankan distribusi aslinya. Hal ini bertujuan untuk memastikan bahwa evaluasi performa model dilakukan secara objektif dan mencerminkan kondisi ketidakseimbangan yang sesungguhkan pada data asli.

Setelah dilakukan proses pemodelan, dilakukan pula prediksi anomali berdasarkan dua fitur yang telah dinormalisasikan, yaitu *scaled_Amount* dan *scaled_Time*. Visualisasi ini bertujuan untuk memberikan gambaran dari prediksi klasifikasi antara transaksi normal (label 0) dan transaksi penipuan (label 1) berdasarkan hasil masing-masing model.



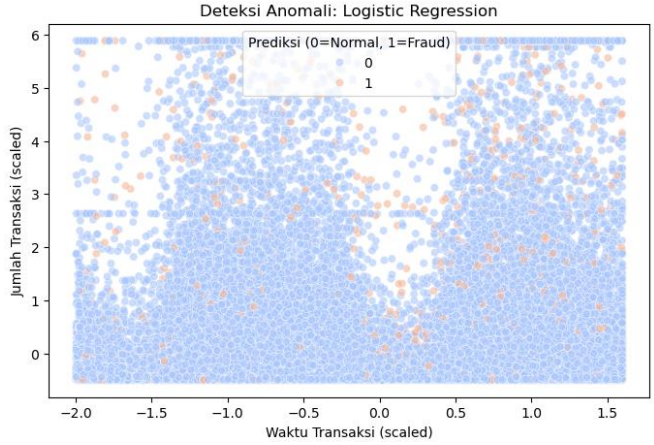
Gambar 4. *Heatmap* Deteksi Anomali *Random Forest*

Pada Gambar 4 terlihat bahwa *Random Forest* mampu mengidentifikasi sebagian besar titik anomali secara tersebar di berbagai waktu dan jumlah transaksi. Titik-titik berwarna jingga merepresentasikan prediksi *fraud* yang tersebar, namun relatif sedikit dibandingkan jumlah total titik. Hal ini menunjukkan bahwa model cukup selektif dalam mengklasifikasikan anomali dan cenderung menghindari *false positive*.



Gambar 5. *Heatmap* Deteksi Anomali XGBoost

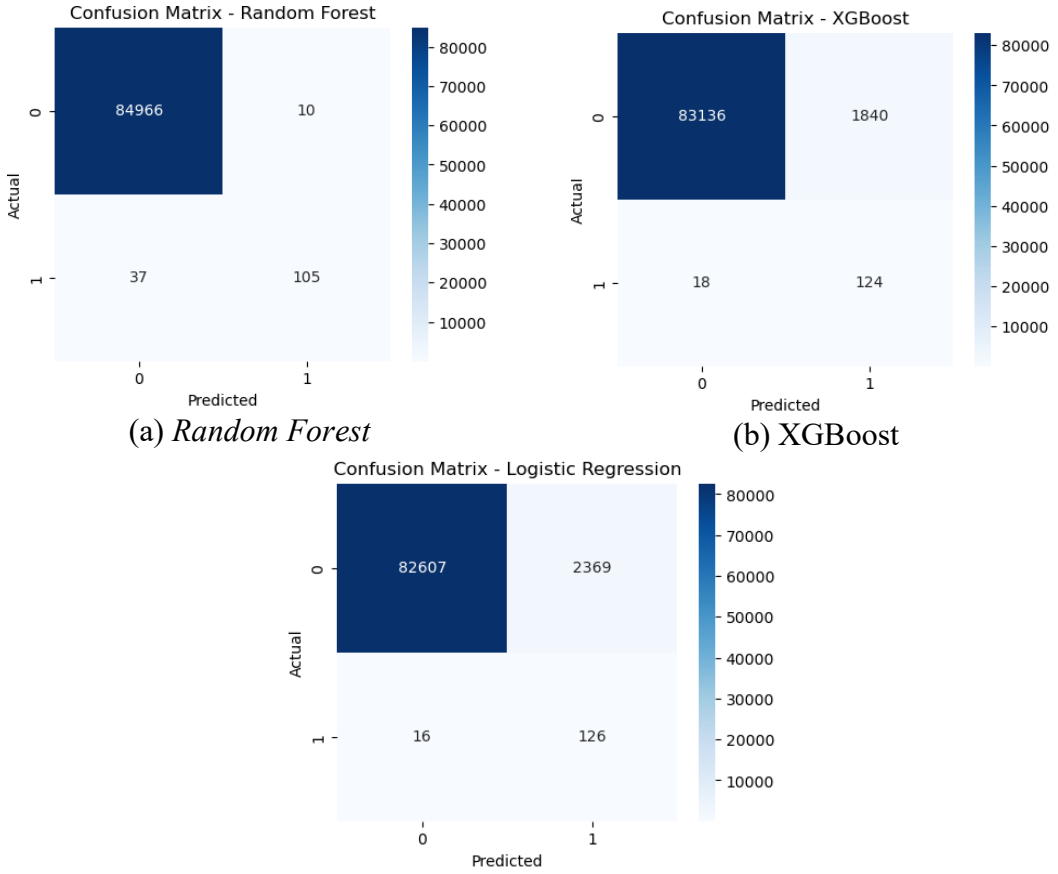
Pada Gambar 5 menampilkan hasil dari XGBoost. Model ini menunjukkan pola deteksi yang sedikit lebih padat pada area tertentu, terutama di sekitar nilai *scaled_Time* mendekati nol. Hal ini mengindikasikan bahwa XGBoost berhasil mengenali lebih banyak pola transaksi penipuan, meskipun dengan risiko menghasilkan lebih banyak *false positive* jika dibandingkan dengan *Random Forest*.



Gambar 6. *Heatmap* Deteksi Anomali *Logistic Regression*

Sementara itu, Gambar 6 yang merepresentasikan *Logistic Regression* menunjukkan sebaran prediksi anomali yang lebih banyak, dengan banyaknya titik *fraud* yang tersebar di seluruh rentang waktu dan jumlah transaksi. Pola ini menandakan bahwa model cenderung *overpredict* terhadap kelas minoritas (*fraud*), sehingga menghasilkan *recall* yang tinggi namun memiliki nilai *precision* yang rendah.

Setelah model dilatih dengan menggunakan data hasil *balancing* SMOTE dan tiga model klasifikasi yaitu *Logistic Regression*, *Random Forest* dan XGBoost (yang telah dioptimasi *hyperparameter* dengan menggunakan Optuna). Fokus utama evaluasi model dalam kemampuannya mendeteksi transaksi penipuan (kelas minoritas) dengan memperhatikan nilai metrik evaluasi utama, yaitu *precision, recall, f1-score,* serta ROC AUC. Berikut merupakan perbandingan *confusion matrix* dari ketiga model Gambar 7(a) untuk model *Random Forest*, Gambar 7(b) untuk model XGBoost, dan Gambar 7(c) untuk model *Logistic Regression*.



(a) *Random Forest* (b) XGBoost

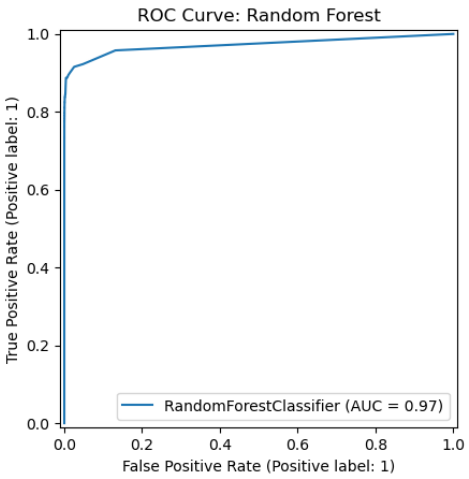
(c) *Logistic Regression*

Gambar 7. Perbandingan *Confusion Matrix* Model

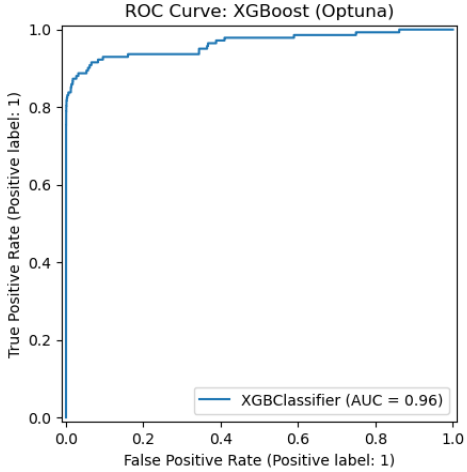
Berdasarkan *confusion matrix* ketiga model klasifikasi, dapat dijelaskan bahwa algoritma XGBoost dan *Logistic Regression* pada Gambar 7 (b) dan (c) menunjukkan tingkat deteksi penipuan yang lebih tinggi, masing-masing dengan 124 dan 126 transaksi yang berhasil diidentifikasi. Namun, keduanya menghasilkan jumlah *false positive* yang jauh lebih besar, yaitu 1840 untuk XGBoost dan 2369 untuk *Logistic Regression*. Hal ini mengindikasikan bahwa kedua model tersebut cenderung lebih agresif dalam memprediksi penipuan, yang berdampak pada meningkatnya kemungkinan terjadinya *false alarm* dalam sistem operasional. Akibatnya, model ini dapat menyebabkan ketidaknyamanan bagi pengguna kartu kredit yang sah karena kesalahan pemblokiran transaksi.

Berbeda dengan *random forest* pada Gambar 7 (a) menunjukkan performa yang paling seimbang dalam mendeteksi transaksi penipuan kartu kredit. Model *Random Forest* berhasil mengklasifikasikan 84.966 transaksi normal dengan benar (*true negative*) dan hanya menghasilkan 10 kesalahan klasifikasi sebagai penipuan (*false positive*). Selain itu, model *Random Forest* mampu mendeteksi 105 transaksi penipuan (*true positive*) dan gagal mendeteksi sebanyak 37 transaksi (*false negative*). Keseimbangan antara *precisien* dan *recall* pada *Random Forest* menjadikan menjadikan model efektif dan efisien untuk digunakan dalam konteks sistem deteksi penipuan yang sensitif terhadap kesalahan klasifikasi.

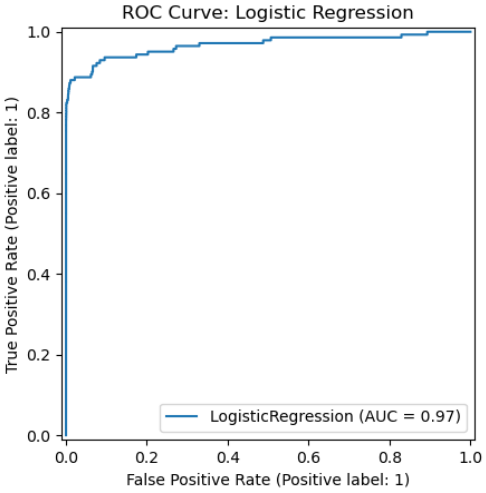
Berdasarkan Gambar 8(a) menunjukkan kurva ROC untuk model *Random Forest*, Gambar 8(b) untuk model XGBoost, dan Gambar 8(c) untuk model *Logistic Regression*.



(a) *Random Forest*



(b) XGBoost



(c) *Logistic Regression*

Gambar 8. Perbandingan ROC-AUC Curve model klasifikasi

Kurva ROC (*Receiver Operating Characteristic*) merupakan grafis yang banyak digunakan dalam menilai kinerja model klasifikasi biner. Kurva ini menggambarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR). Selain itu, nilai *Area Under Curve* (AUC) digunakan untuk mengukur luas di bawah kurva tersebut dimana semakin mendekati angka 1 maka semakin bagus performa model. Berdasarkan Gambar 8 hasil visualiasi kurva ROC untuk tiga model yang telah dibangun dan diuji yang masing-masing model memiliki nilai AUC sebagai berikut.

- XGBoost : AUC = 0,96.
- *Logistic Regression* : AUC = 0,97.
- *Random Forest* : AUC = 0,96.

Ketiga model memiliki performa yang sangat baik dalam membedakan transaksi penipuan (*fraud*) dan transaksi normal (*non-fraud*). Pada grafik *Logistic Regression* dan *Random Forest* menunjukkan bahwa kurva naik ke arah sudut kiri atas, yang mendefinisikan bahwa kedua model tersebut memiliki tingkat kesalahan positif yang rendah serta tingkat keberhasilan dalam mendeteksi yang tinggi. Hal tersebut dapat didefinisikan kemampuannya yang kuat dalam mengidentifikasi kasus penipuan. Sementara itu, pada grafik XGBoost juga memiliki performa yang baik meskipun kurva sedikit turun pada sudut kiri atas dan bagian tengah kurva. Hal ini mendefinisikan bahwa XGBoost lebih cenderung

menghasilkan prediksi positif paslu (*false positive*) pada tingkat ambang tertentu, meskipun secara keseluruhan model XGBoost masih memberikan hasil klasifikasi yang baik.

Hasil evaluasi juga dapat dilihat dengan memperhatikan metrik evaluasi, yaitu *precision, recall,* dan *f1-score*. Untuk perbandingan hasil metrik evaluasi dapat dilihat berdasarkan tabel berikut.

Tabel 1. Perbandingan hasil evaluasi model klasifikasi

Model	*Precision (Fraud)*	*Recall (Fraud)*	*F1-Score (Fraud)*	ROC AUC
Random Forest	0.91	0.74	0.82	0.8697
XGBoost (Optuna)	0.07	0.87	0.14	0.9275
Logistic Regression	0.05	0.89	0.10	0.9297

Berdasarkan Tabel 1 dapat dilihat performa dari ketiga model klasifikasi dalam menangani kasus penipuan. Model *logistic regression* memiliki nilai *recall* tertinggi dibandingkan dua model lainnya yaitu sebesar 0,89. Namun memiliki nilai *precision* yang rendah dibandingkan kedua model lainnya yaitu sebesar 0,05. Berdasarkan nilai tersebut model *logistic regression* sangat sensitif terhadap deteksi *fraud*, namun banyak memberikan *false positive*, sehingga tidak efektif dalam prkatik karena dapat menimbulkan kesalahan pemblokiran transaksi yang sah. Namun sebaliknya, model *random forest* menunjukkan performa paling seimbang dibandingkan dua model yang lainnya, dengan nilai *precision* sebesar 0,91 dan *recall* 0,74 serta *f1-score* tertinggi sebesar 0,82 yang menunjukkan bahwa model *random forest* tidak hanya sensitif tetai juga cukup selektif dalam mengidentifikasi *farud*. Hal ini dapat menjadikan *random forest* sebagai model dengan *trade-off* terbaik antara deteksi *fraud* dan kesalahan klasifikasi.

Adapun XGBoost (Optuna) yang menghasilkan niai ROC-AUC tertinggi dibandingkan kedua model klasifikasi lainnya namun memiliki nilai *precison* yang masih rendah yaitu sebesar 0,07. Hal ini menunjukkan bahwa meskipun model mampu mebedakan kelas dengan baik secara probabiliastik, namun klasifikasi *farud* yang benar tetap sedikit jika dibandingkan dengan keseluruhan prediksi *fraud*.

5. KESIMPULAN

Penelitian menggunakan tiga algoritma *machine learning*, yaitu *Logistic Regression*, *Random Forest*, dan XGBoost. Penelitian ini dilakukan dengan tujuan merancang dan mengevaluasi mode deteksi penipuan kartu kredit dengan melakukan penanganan data yang tidak seimbang menggunakan teknik SMOTE dan *tuning hyperparameter* dilakukan pada XGBoost dengan Optuna. Hasil evaluasi menunjukkan bahwa *Random Forest* merupakan model terbaik, dengan *precision* sebesar 0,91, *recall* 0,74, dan f1-*score* 0,82 pada kelas penipuan. *Logistic Regression* memiliki *recall* tertinggi (0,89) namun *precision* sangat rendah (0,05), sedangkan XGBoost menampilkan AUC yang tinggi (0,96) tetapi kurang optimal pada f1-*score*.

Secara keseluruhan, *Random Forest* dinilai paling seimbang dan efektif dalam mendeteksi transaksi penipuan karena mampu menjaga akurasi sekaligus meminimalkan kesalahan klasifikasi. Model ini tidak hanya akurat dari sisi metrik, tetapi juga memberikan kestabilan performa dan interpretabilitas yang memadai untuk diterapkan dalam sistem pemantauan transaksi di dunia nyata. Penelitian selanjutnya disarankan untuk menguji model pada data *real-time* dan menambahkan fitur perilaku pengguna guna meningkatkan akurasi. Selain itu, penggunaan model *deep learning* seperti LSTM atau *autoencoder* dapat dieksplorasi untuk hasil yang lebih adaptif dan interpretatif.

DAFTAR PUSTAKA

Akiba, Takuya, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. "Optuna: A Next-Generation Hyperparameter Optimization Framework." *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: 2623–31. doi:10.1145/3292500.3330701.

Alamri, Maram, and Mourad Ykhlef. 2024. "Hybrid Feature Engineering Based on Customer Spending Behavior for Credit Card Anomaly and Fraud Detection." *Electronics (Switzerland)* 13(20). doi:10.3390/electronics13203978.

Carcillo, F., Dal Pozzolo, A. Dal Pozzolo, Y.-A Le Borgne, O. Caelan, Y Mazzer, and G. Bontempi. 2019. "SCARFF: A Scalable Framework for Streaming Credit Card Fraud Detection with Spark." *Information Fusion*: 297–308.

Carcillo, Fabrizio, Yann-Ael Le Borgne, Olivier Caelan, Yacine Kessaci, Oble Frederic, and Gianluca Bontempi. 2021. "Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection." *Information Sciences* 557: 317–31.

Chawla, Nitesh V, Kevin W Bowyer, and Lawrence O Hall. 2002. "SMOTE : Synthetic Minority Over-Sampling Technique." 16: 321–57.

Correa Bahnsen, Alejandro, Djamila Aouada, Aleksandar Stojanovic, and Björn Ottersten. 2016. "Feature Engineering Strategies for Credit Card Fraud Detection." *Expert Systems with Applications* 51: 134–42. doi:10.1016/j.eswa.2015.12.030.

Hosmer, David W., Stanley Lemeshow, and Rodney X. Sturdivant. 2013. *Applied Logistic Regression*.

Jiang, Shanshan, Ruiting Dong, Jie Wang, and Min Xia. 2023. "Credit Card Fraud Detection Based on Unsupervised Attentional Anomaly Detection Network." *Systems* 11(6): 1–14. doi:10.3390/systems11060305.

Kotsiantis, Sotiris, and Dimitris Kanellopoulos. 2006. "Association Rules Mining: A Recent Overview." *Science* 32(1): 71–82.

Mbakwe, Amarachi Blessing, and Sikiru Ademola Adewale. 2022. "MACHINE LEARNING ALGORITHMS FOR CREDIT CARD FRAUD DETECTION Amarachi." *International Journal of Computer Science, Engineering and Applications* 12(3/4/5/6): 01–13. doi:10.5121/mlaij.2022.9402.

Purwa, Taly. 2019. "Perbandingan Metode Regresi Logistik Dan Random Forest Untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin Di Kabupaten Karangasem, Bali Tahun 2017)." *Jurnal Matematika, Statistika dan Komputasi* 16(1): 58. doi:10.20956/jmsk.v16i1.6494.

Ronny Susetyoko, Wiratmoko Yuwono, and Elly Purwantini. 2022. "Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP Kuliah Menggunakan Regresi Logistik Biner." *Jurnal Informatika Polinema* 8(4): 31–40. doi:10.33795/jip.v8i4.914.

Science, Computer, Varun Chellapilla, Sravya Chikkam, and Jayanth Sriram Melinati. 2024. "CREDIT CARD FRAUD DETECTION USING A STACKING ENSEMBLE APPROACH WITH LSTM AND RANDOM FOREST." (April): 16–19.

Syeeda, P., and A. Mohanty. 2024. "Fraud Detection in Financial Transactions Using Machine Learning." *International Journal of Scientific Research & Engineering Trends* 10(5).

Witten, I. H., Eibe. Frank, Mark A. Hall, and Christopher J. Pal. 2017. Data mining - Practical machine learning tools and techniques *CHAPTER 10. Deep Learning*.