



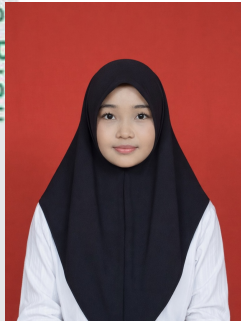
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

MODEL DETEKSI PENIPUAN TRANSAKSI KARTU KREDIT BERBASIS *BOOSTING CLASSIFIERS* DENGAN OPTIMASI *HYPERPARAMETER* DAN *EXPLAINABLE AI* (XAI)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer pada
Program Studi Sistem Informasi



Oleh:

AMELIA RAHMADANI
12250320334



UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2026

LEMBAR PERSETUJUAN
MODEL DETEKSI PENIPUAN TRANSAKSI KARTU KREDIT
BERBASIS *BOOSTING CLASSIFIERS* DENGAN OPTIMASI
HYPERPARAMETER* DAN *EXPLAINABLE AI (XAI)

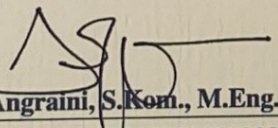
TUGAS AKHIR

Oleh:

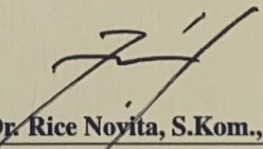
AMELIA RAHMADANI
12250320334

Telah diperiksa dan disetujui sebagai laporan tugas akhir
di Pekanbaru, pada tanggal 20 Januari 2026

Ketua Program Studi


Angraini, S.Kom., M.Eng., Ph.D.
NIP. 198408212009012008

Pembimbing


Dr. Rice Noyita, S.Kom., M.Kom.
NIP. 198511272023212032

LEMBAR PENGESAHAN
MODEL DETEKSI PENIPUAN TRANSAKSI KARTU KREDIT
BERBASIS *BOOSTING CLASSIFIERS* DENGAN OPTIMASI
HYPERPARAMETER* DAN *EXPLAINABLE AI (XAI)

TUGAS AKHIR

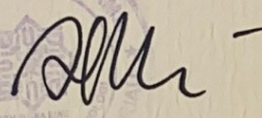
Oleh:

AMELIA RAHMADANI
12250320334

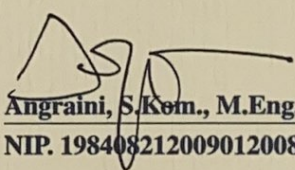
Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 8 Januari 2026

Pekanbaru, 8 Januari 2026
Mengesahkan,

Dekan


Dr. Yuslenita Muda, S.Si., M.Sc.
NIP. 197701032007102001

Ketua Program Studi


Angraini, S.Kom., M.Eng., Ph.D.
NIP. 198408212009012008

DEWAN PENGUJI:

Ketua : Muhammad Jazman, S.Kom., M.Infosys.

Sekretaris : Dr. Rice Novita, S.Kom., M.Kom.

Anggota 1 : Mustakim, ST., M.Kom.

Anggota 2 : Eki Saputra, S.Kom., M.Kom.

Lampiran Surat

Nomor : Nomor 25/2021

Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Amelia Rahmadani

NIM : 12250320334

Tempat/Tgl. Lahir : Dumai / 09 November 2003

Fakultas/Pascasarjana : Fakultas Sains dan Teknologi

Prodi : Sistem Informasi

Judul Skripsi :

MODEL DETEKSI PENIPUAN TRANSAKSI KARTU KREDIT BERBASIS *BOOSTING*
CLASSIFIERS DENGAN OPTIMASI *HYPERPARAMETER* DAN *EXPLAINABLE AI (XAI)*

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan Skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu Skripsi saya ini, saya nyatakan bebas dari plagiat.
4. Apa bila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 20 Januari 2026

Yang membuat pernyataan


METERAI
TEMPEL
829B2ALX351455109

AMELIA RAHMADANI

NIM. 12250320334



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum, dengan ketentuan bahwa hak cipta ada pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan atas izin penulis dan harus dilakukan mengikuti kaedah dan kebiasaan ilmiah serta menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin tertulis dari Dekan fakultas universitas. Perpustakaan dapat meminjamkan Tugas Akhir ini untuk anggotanya dengan mengisi nama, tanda peminjaman dan tanggal pinjam pada *form* peminjaman.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



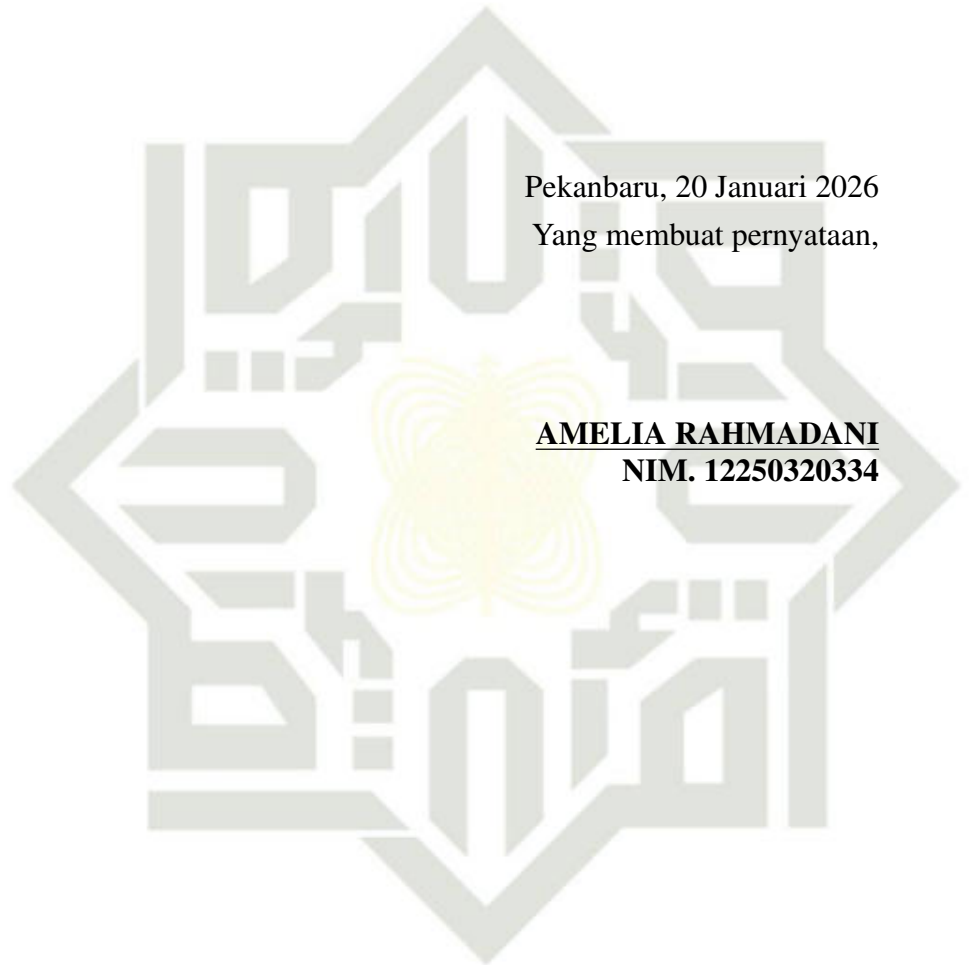


LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 20 Januari 2026
Yang membuat pernyataan,

AMELIA RAHMADANI
NIM. 12250320334



UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillah wa syukurillah, sebagai bentuk rasa syukur peneliti atas segala nikmat yang telah didapatkan tanpa ada kekurangan sekalipun. Sholawat beserta salam tak lupa peneliti ucapkan kepada Nabi Muhammad Shalallahu 'Alaihi Wa Sallam dengan melafalkan *Allahumma Shalli 'ala Sayyidina Muhammad wa 'ala Ali Sayyidina Muhammad*. Semoga kita semua senantiasa mendapat syafaat Nya di dunia maupun di akhirat. *Amiin Ya Rabbal 'Alamiin*. Tugas Akhir ini peneliti persembahkan kepada Ayahanda Denny Dharma Putra dan Ibunda Desi Arisandi yang tidak henti-hentinya memberikan semangat, dukungan, dan doa kepada peneliti sehingga dapat menyelesaikan pendidikan Strata Satu (S-1) dengan baik. Akhirnya peneliti berhasil menyelesaikan Laporan Tugas Akhir ini. Tugas Akhir ini juga peneliti persembahkan kepada diri peneliti sendiri, yang telah dengan sabar mengusahakan diri untuk tetap bertahan dan terus melangkah. Terima kasih karena telah melalui setiap proses dengan segala rasa yang menyertainya, tetap berusaha berdiri kembali, serta memilih untuk tetap percaya pada diri sendiri meskipun keraguan sempat hadir. Setiap langkah yang ditempuh menjadi bagian berharga dari proses pendewasaan diri hingga akhirnya sampai pada titik ini. Terima kasih kepada Ibu Dr. Rice Novita S. Kom., M. Kom. yang telah sabar membimbing peneliti untuk menyelesaikan Tugas Akhir ini dan selalu menyempatkan waktu untuk mengajari serta membimbing peneliti dalam semua hal yang berkaitan dengan Tugas Akhir ini. Semoga Allah Subhanahu Wa Ta'ala memberikan keberkahan serta rahmat Nya dan senantiasa menjadikan kita hamba yang terus berdoa kepada-Nya. *Amiin Ya Rabbal 'Alamiin*.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.



KATA PENGANTAR

Alhamdulillah Rabbil 'Aalamiin. Segala puji bagi Allah SWT atas limpahan rahmat, taufik, dan karunia-Nya sehingga peneliti dapat menyelesaikan Tugas Akhir ini dengan baik. Shalawat serta salam senantiasa peneliti haturkan kepada Nabi Muhammad SAW, sebagai suri teladan dalam menuntut ilmu dan menjaga amanah, dengan ucapan “*Allahumma shalli ‘ala Sayyidina Muhammad wa ‘ala ali Sayyidina Muhammad*”.

Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau. Adapun judul penelitian ini adalah **“Model Deteksi Penipuan Transaksi Kartu Kredit Berbasis Boosting Classifiers dengan Optimasi Hyperparameter dan Explainable AI (XAI).”** Penelitian ini lahir dari perhatian peneliti terhadap meningkatnya kompleksitas risiko penipuan transaksi pada ekosistem keuangan digital, sekaligus keinginan untuk menghadirkan pendekatan yang tidak hanya kuat secara performa, tetapi juga dapat dipahami, dan relevan untuk kebutuhan pengambilan keputusan.

Dalam proses penyusunan penelitian ini, peneliti menyadari bahwa perjalanan menyelesaikan Tugas Akhir bukan semata kegiatan akademik, melainkan proses belajar yang membentuk ketekunan, disiplin, serta keberanian untuk menguji asumsi dan memperbaiki kekurangan. Peneliti memperoleh banyak arahan, dukungan, dan doa dari berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini peneliti menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Ibu Prof. Dr. Hj. Leny Nofianti MS, SE., M.Si., Ak., CA sebagai Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Ibu Dr. Yuslenita Muda, S.Si., M.Sc sebagai Dekan Fakultas Sains dan Teknologi.
3. Ibu Angraini, S.Kom., M.Eng., Ph.D sebagai Ketua Program Studi Sistem Informasi.
4. Ibu Dr. Rice Novita, S.Kom., M.Kom sebagai Dosen Pembimbing yang dengan sabar mendampingi peneliti selama proses penyusunan Tugas Akhir ini. peneliti sangat terbantu oleh arahan yang jelas, koreksi yang teliti, dan dorongan untuk terus memperbaiki bagian-bagian yang kurang. Bimbingan tersebut membuat peneliti tidak hanya menyelesaikan penelitian ini, tetapi juga belajar membangun cara berpikir yang lebih rapi, kritis, dan bertanggung jawab dalam menulis Tugas Akhir.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5. Yang tercinta, Mamak. terima kasih yang rasanya tidak akan pernah cukup untuk menggambarkan semuanya. Terima kasih untuk doa yang tidak pernah putus, untuk sabar yang tidak pernah habis, dan untuk kasih sayang yang selalu jadi tempat pulang. Saat peneliti lelah, ragu, bahkan merasa tidak sanggup, mamak tetap ada menenangkan, menguatkan, dan membuat peneliti percaya bahwa semua proses ini bisa dilewati. Peneliti paham, banyak hal yang mamak korbankan diam-diam. Semoga mamak selalu diberi kesehatan dan umur yang panjang, supaya peneliti punya kesempatan membalas semua kebaikan itu, satu per satu, dengan cara terbaik.
6. Yang tercinta, Ayah. Terima kasih atas doa, dukungan, dan keteguhan yang selalu ayah berikan. Peneliti mungkin tidak selalu pandai mengungkapkan, tapi peneliti benar-benar merasa bahwa ayah selalu ada di belakang peneliti mendoakan, menguatkan, dan mempercayai peneliti. Kepercayaan itu jadi salah satu alasan peneliti berusaha bertahan sampai selesai.
7. Saudara peneliti, Rahmad Hidayat dan Abimanyu Dharma Wangsa, terima kasih atas dukungan, perhatian, dan doa yang selalu kalian berikan kepada peneliti. Kebersamaan, nasihat, dan semangat yang kalian hadirkan menjadi penguat dalam banyak proses, terutama ketika peneliti sedang lelah atau merasa ragu. Dukungan kalian membantu terus menjaga langkah peneliti selama menempuh pendidikan hingga Tugas Akhir ini dapat diselesaikan.
8. Sepupu peneliti, Nursyahira. Terima kasih karena sudah menjadi bagian penting dalam proses ini. Meski tidak selalu berada dalam satu waktu dan ruang yang sama, dukunganmu tetap terasa. Terima kasih karena menjadi keluarga sekaligus teman yang mau mendengar, memahami, dan berpihak pada peneliti. Semoga ke depan ada lebih banyak hari yang ringan, dan lebih banyak tawa yang bisa kita bagi bersama.
9. Teman-teman peneliti, Riski Permata Dini, Raudah Islamiah, Rifka Anrahvi, Stevani, dan Saumi Sintia, terima kasih atas kebersamaan, dukungan, dan perhatian yang diberikan selama masa perkuliahan hingga proses penyusunan Tugas Akhir ini. Kehadiran kalian dalam berbagai bentuk sederhana membantu peneliti melewati hari-hari yang melelahkan, menjaga semangat, dan membuat proses ini terasa lebih ringan untuk dijalani.
10. Terakhir, untuk diri peneliti sendiri. Terima kasih karena sudah bertahan sampai sejauh ini. Terima kasih sudah tetap berjalan meski tidak selalu yakin, tetap berusaha meski sering lelah, dan tetap menyelesaikan apa yang sudah dimulai. Terima kasih juga karena terus berusaha menjaga hati tetap lembut di tengah hari-hari yang kadang membuat segalanya terasa cepat,



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

bising, dan melelahkan. Semoga pencapaian ini menjadi pengingat bahwa peneliti mampu melewati proses yang berat dengan cara yang baik, dan tetap menjadi pribadi yang tenang serta bertumbuh.

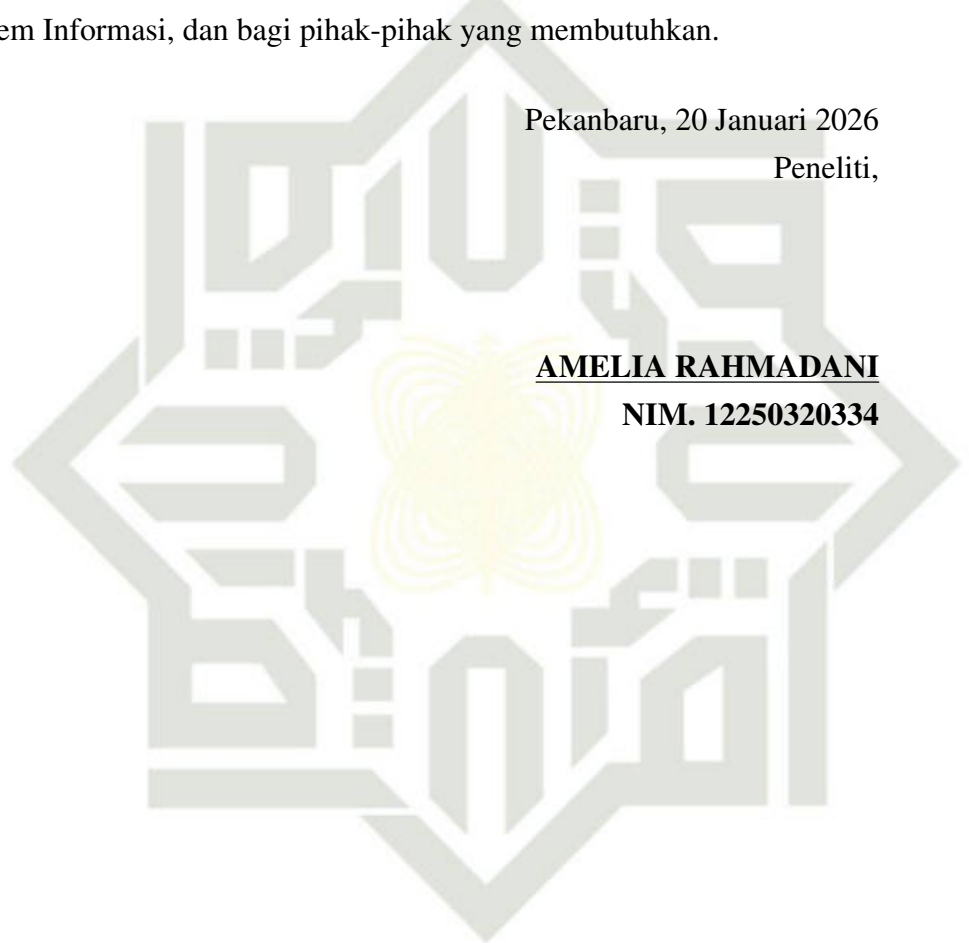
Akhir kata, peneliti menyadari bahwa Tugas Akhir ini masih memiliki keterbatasan, baik dari segi penulisan maupun pembahasan. Oleh karena itu, peneliti mengharapkan kritik dan saran yang bersifat membangun demi penyempurnaan penelitian ini di masa mendatang. Semoga Tugas Akhir ini dapat memberikan manfaat serta kontribusi yang positif bagi pengembangan ilmu pengetahuan, khususnya di bidang Sistem Informasi, dan bagi pihak-pihak yang membutuhkan.

Pekanbaru, 20 Januari 2026

Peneliti,

AMELIA RAHMADANI

NIM. 12250320334



UIN SUSKA RIAU



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

MODEL DETEKSI PENIPUAN TRANSAKSI KARTU KREDIT BERBASIS *BOOSTING CLASSIFIERS* DENGAN OPTIMASI *HYPERPARAMETER* DAN *EXPLAINABLE AI* (XAI)

AMELIA RAHMADANI

NIM: 12250320334

Tanggal Sidang: 8 Januari 2026

Periode Wisuda:

Program Studi Sistem Informasi

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl. Soebrantas, No. 155, Pekanbaru

ABSTRAK

Peningkatan volume transaksi digital menyebabkan institusi keuangan menghadapi risiko penipuan kartu kredit yang semakin kompleks, dengan karakteristik data yang sangat tidak seimbang. Penelitian ini bertujuan membandingkan kinerja tiga algoritma *boosting* modern, yaitu *XGBoost*, *LightGBM*, dan *CatBoost* pada skenario data sangat tidak seimbang, sekaligus mengkaji peran *Synthetic Minority Over-sampling Technique* (SMOTE), *hyperparameter tuning* GridSearchCV, dan *explainable artificial intelligence* (XAI) berbasis *SHapley Additive exPlanations* (SHAP). Dataset yang digunakan merupakan data sintesis transaksi kartu kredit dari Kaggle yang berisi 1.852.394 juta transaksi dengan *fraud rate* sekitar 0,52%. Ketidakseimbangan kelas ditangani menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) dengan rasio 10% pada data latih, diikuti *hyperparameter tuning* berbasis GridSearchCV dan penyesuaian *decision threshold* berbasis *F1-score*. Hasil menunjukkan bahwa seluruh model mencapai AUC-ROC di atas 0,996 dan *F1-score* sekitar 0,90 pada data uji. CatBoost menghasilkan kombinasi terbaik antara *F1-score* dan Recall, sedangkan XGBoost mencapai *Precision* tertinggi. Analisis SHAP mengonfirmasi bahwa model bergantung pada kombinasi fitur jumlah transaksi, pola waktu transaksi, perilaku historis kartu, dan tingkat risiko *merchant* dalam keputusan klasifikasi *fraud*, sehingga meningkatkan transparansi dan akuntabilitas sistem deteksi penipuan yang dihasilkan.

Kata Kunci: Deteksi penipuan kartu kredit, *XGBoost*, *LightGBM*, *CatBoost*, kelas tidak seimbang, SMOTE, *hyperparameter tuning*, SHAP.



COMPARISON OF CREDIT CARD TRANSACTION FRAUD DETECTION MODELS BASED ON BOOSTING CLASSIFIERS WITH HYPERPARAMETER OPTIMIZATION AND EXPLAINABLE AI (XAI)

AMELIA RAHMADANI
NIM: 12250320334

Date of Final Exam: January 8th 2026
Graduation Period:

Department of Information System
Faculty of Science and Technology
State Islamic University of Sultan Syarif Kasim Riau
Soebrantas Street, No. 155, Pekanbaru

ABSTRACT

The increase in digital transaction volume has caused financial institutions to face increasingly complex credit card fraud risks, with highly imbalanced data characteristics. This study aims to compare the performance of three modern boosting algorithms, namely XGBoost, LightGBM, and CatBoost in a highly imbalanced data scenario, while also examining the role of Synthetic Minority Over-sampling Technique (SMOTE), hyperparameter tuning GridSearchCV, and explainable artificial intelligence (XAI) based on SHapley Additive exPlanations (SHAP). The dataset used is synthetic credit card transaction data from Kaggle, containing 1,852,394 million transactions with a fraud rate of approximately 0.52%. Class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE) with a ratio of 10% on the training data, followed by hyperparameter tuning based on GridSearchCV and decision threshold adjustment based on F1-score. The results show that all models achieved an AUC-ROC above 0.996 and an F1-score of around 0.90 on the test data. CatBoost produced the best combination of F1-score and Recall, while XGBoost achieved the highest Precision. SHAP analysis confirmed that the model relied on a combination of transaction amount, transaction time patterns, historical card behavior, and merchant risk level in fraud classification decisions, thereby improving the transparency and accountability of the resulting fraud detection system.

Keywords: Credit card fraud detection, XGBoost, LightGBM, CatBoost, imbalanced classes, SMOTE, hyperparameter tuning, SHAP.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
KATA PENGANTAR	vii
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI	xii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvi
DAFTAR SINGKATAN	xvii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan	5
1.5 Manfaat	5
1.6 Sistematika Penulisan	6
LANDASAN TEORI	7
2.1 <i>Machine Learning</i>	7
2.1.1 <i>Supervised Learning</i>	7
2.1.2 <i>Unsupervised Learning</i>	7
2.1.3 <i>Semi-Supervised Learning</i>	7
2.1.4 <i>Reinforcement Learning</i>	8
2.2 <i>Categorical Boosting (CatBoost)</i>	8



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.3	<i>Extreme Gradient Boosting (XGBoost)</i>	9
2.4	<i>Light Gradient Boosting Machine (LightGBM)</i>	11
2.5	<i>Synthetic Minority Over-Sampling Technique (SMOTE)</i>	12
2.6	<i>Hyperparameter Tuning dengan Grid Search Cross Validation</i>	13
2.7	<i>Explainable Artificial Intelligent (XAI)</i>	13
2.8	<i>Shapley Additive Explanations (SHAP)</i>	15
2.9	<i>Confusion Matrix</i>	17
2.10	<i>Area Under the Curve</i>	18
2.11	Deteksi Penipuan Kartu Kredit	18
2.12	<i>Google Colab</i>	21
2.13	Penelitian Terdahulu	21
3	METODOLOGI PENELITIAN	25
3.1	Metodologi Penelitian	25
3.1.1	Tahap Perencanaan	26
3.1.1.1	Identifikasi Masalah	26
3.1.1.2	Menentukan Tujuan	26
3.1.1.3	Studi Literatur	26
3.1.2	Pengumpulan Data	26
3.1.3	Validasi Data	27
3.1.4	<i>Preprocessing Data</i>	28
3.1.4.1	<i>Exploratory Data Analysis (EDA)</i>	28
3.1.4.2	<i>Cleaning Data</i>	28
3.1.4.3	<i>Feature Engineering</i>	29
3.1.4.4	<i>Train-Test Split Data</i>	29
3.1.4.5	<i>Imputasi dan One-Hot Encoding</i>	30
3.1.5	<i>Oversampling Data</i>	30
3.1.6	Pemodelan <i>Machine Learning</i>	30
3.1.7	<i>Hyperparameter Tuning</i>	31
3.1.8	Evaluasi Model <i>Machine Learning</i>	32
3.1.8.1	Analisis Interpretabilitas dengan SHAP	33
4	ANALISA DAN HASIL	34
4.1	Tahap Perencanaan	34
4.1.1	Identifikasi Masalah	34
4.1.2	Menentukan Tujuan	35
4.1.3	Studi Pustaka	35
4.2	Proses Pengumpulan Data	35

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.3	<i>Exploratory Data Analysis (EDA)</i>	39
4.4	<i>Data Preprocessing</i>	42
4.4.1	<i>Cleaning Data</i>	42
4.4.2	<i>Feature Engineering</i>	43
4.4.3	<i>Train-Test-Validation Split Data</i>	45
4.4.4	<i>Imputasi dan One-Hot Encoding</i>	46
4.5	<i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	46
4.6	<i>Pemodelan Machine Learning</i>	48
4.6.1	<i>Pemodelan Machine Learning Tanpa SMOTE</i>	48
4.6.2	<i>Pemodelan Machine Learning dengan SMOTE</i>	49
4.7	<i>Hyperparameter Tuning dengan GridSearchCV</i>	51
4.7.1	<i>Skema GridSearchCV</i>	51
4.7.2	<i>Pemilihan Kombinasi Hyperparameter Terbaik</i>	51
4.7.3	<i>Penyesuaian Threshold Probabilitas Kelas Fraud</i>	52
4.7.4	<i>Evaluasi Model pada Data Validasi</i>	53
4.8	<i>Evaluasi Model Terbaik pada Data Uji</i>	54
4.9	<i>Analisis Interpretabilitas dengan SHAP</i>	56
4.9.1	<i>Interpretabilitas Model XGBoost</i>	57
4.9.2	<i>Interpretabilitas Model LightGBM</i>	58
4.9.3	<i>Interpretabilitas Model CatBoost</i>	59
4.10	<i>Pembahasan Hasil</i>	60
5	PENUTUP	62
5.1	<i>Kesimpulan</i>	62
5.2	<i>Saran</i>	63

DAFTAR PUSTAKA

LAMPIRAN A SURAT PERNYATAAN VALIDASI PAKAR A - 1

LAMPIRAN B DOKUMENTASI B - 1

DAFTAR GAMBAR

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.1	Struktur CatBoost	8
2.2	Struktur XGBoost	10
2.3	Struktur LightGBM	11
2.4	Visualisasi <i>Oversampling</i>	12
2.5	Penggunaan XAI di ML	14
2.6	Klasifikasi metode Explainable AI (XAI)	14
3.1	Metodologi Penelitian	25
4.1	Distribusi Transaksi Normal dan <i>Fraud</i>	39
4.2	Pola volume berdasarkan jam transaksi	40
4.3	Pola volume berdasarkan hari dalam seminggu	41
4.4	Pola volume berdasarkan kelompok usia	42
4.5	Pola volume berdasarkan gender	42
4.6	Distribusi Kelas Setelah SMOTE	47
4.7	Plot ringkasan SHAP pengaruh fitur pada prediksi XGBoost	57
4.8	Plot ringkasan SHAP pengaruh fitur pada prediksi LightGBM	58
4.9	Plot ringkasan SHAP pengaruh fitur pada prediksi CatBoost	59
A.1	Surat Pernyataan Validasi Pakar	A - 1
A.2	Surat Pernyataan Validasi Pakar	A - 2
B.1	Dokumentasi	B - 1



DAFTAR TABEL

2.1	Confusion Matrix	17
2.2	Penelitian Terdahulu	21
3.1	Nilai Terbaik untuk <i>Hyperparameter</i>	31
4.1	Karakteristik Data	36
4.2	Dataset yang Digunakan	38
4.3	Jarak Antara <i>Customer</i> dan <i>Merchant</i> (km)	40
4.4	Hasil <i>Feature Engineering</i>	43
4.5	Hasil Pembagian Data	45
4.6	<i>Confusion Matrix</i> Tanpa SMOTE	48
4.7	<i>Evaluation Matrix</i> Tanpa SMOTE	48
4.8	Perbandingan Metrik Evaluasi Rasio SMOTE	50
4.9	<i>Confusion Matrix</i> dengan SMOTE	51
4.10	<i>Hyperparameter</i> Terbaik	51
4.11	Hasil <i>Confusion Matrix</i> dengan GridSearchCV + <i>Threshold</i>	53
4.12	<i>Evaluation Matrix</i> dengan GridSearchCV + <i>Threshold</i>	54
4.13	Hasil <i>Confusion Matrix</i> Model Terbaik dengan Data Uji	54
4.14	<i>Evaluation Matrix</i> Model Terbaik dengan Data Uji	55
4.15	Perbandingan Hasil Evaluasi Model	56

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR SINGKATAN

AI	: <i>Artificial Intelligence</i>
AUC	: <i>Area Under the Curve</i>
AUC-ROC	: <i>Area Under the ROC Curve</i>
CatBoost	: <i>Categorical Boosting</i>
EDA	: <i>Exploratory Data Analysis</i>
FN	: <i>False Negative</i>
FP	: <i>False Positive</i>
GBDT	: <i>Gradient Boosting Decision Tree</i>
GridSearchCV	: <i>Grid Search with Cross-Validation</i>
LightGBM	: <i>Light Gradient Boosting Machine</i>
ML	: <i>Machine Learning</i>
NaN	: <i>Not a Number</i>
PCA	: <i>Principal Component Analysis</i>
ROC	: <i>Receiver Operating Characteristic</i>
ROS	: <i>Random Over-Sampling</i>
RUS	: <i>Random Under-Sampling</i>
SHAP	: <i>SHapley Additive exPlanations</i>
SMOTE	: <i>Synthetic Minority Over-sampling Technique</i>
TN	: <i>True Negative</i>
TP	: <i>True Positive</i>
XAI	: <i>Explainable Artificial Intelligence</i>
XGBoost	: <i>eXtreme Gradient Boosting</i>

UIN SUSKA RIAU



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan *machine learning* selama satu dekade terakhir menunjukkan peningkatan yang signifikan dalam kemampuan pemodelan prediktif (Chakraborty, Bhattacharya, Pal, dan Lee, 2024) (Zandieh, Kazemi, dan Ahmadi, 2021). Kemajuan ini didorong oleh peningkatan kapasitas komputasi dan tersedianya data dalam skala besar, memungkinkan penyelesaian masalah klasifikasi dan prediksi yang kompleks melalui berbagai pendekatan algoritmik, mulai dari *supervised learning* untuk klasifikasi transaksi (Afriyie dkk., 2023), dan *deep learning* digunakan secara lebih efektif pada berbagai domain analitik (R. Chen dkk., 2023). Studi oleh (Valencia-Arias dkk., 2025) menekankan bahwa *machine learning* kini telah meluas melintasi berbagai bidang, dengan analisis tren yang memetakan adanya persimpangan lintas disiplin (*interdisciplinary intersections*) yang berkembang pesat untuk menjawab tantangan keberlanjutan global. Di antara berbagai sektor tersebut, sektor keuangan menjadi salah satu arena penerapan yang paling krusial, mengingat institusi finansial menghadapi pola penipuan yang berkembang secara dinamis (Husnaningtyas dan Dewayanto, 2023).

Sektor keuangan mengalami tekanan yang semakin kuat akibat pertumbuhan transaksi digital dan *e-commerce*. Data yang dirilis oleh *Federal Trade Commission (Consumer Sentinel Network Data Book 2021, 2022)* mencatat lebih dari 24 juta kasus penipuan dengan total kerugian mencapai 8,8 miliar dolar, meningkat sekitar 30% dari tahun sebelumnya dan mencakup berbagai bentuk penyalahgunaan akses pembayaran digital, termasuk *fraud* berbasis kartu. Di Indonesia, laporan Survei (*Survei FICO, 2025*) menunjukkan bahwa sekitar 1 dari 4 konsumen di Indonesia (23%) melaporkan mengalami kerugian akibat scam melalui sistem pembayaran *real-time* pada tahun 2024, serta mengalami lonjakan kerugian kategori besar yang melebihi Rp 70 juta. Temuan ini konsisten dengan analisis *Flagright* dalam artikel (*Fraud Detection in Indonesia's Financial Sector, 2025*) yang melaporkan peningkatan *card-not-present fraud*, *identity theft*, dan *social-engineering fraud* seiring pertumbuhan pesat transaksi digital dan *e-commerce* nasional. Selain itu, laporan Tempo dalam artikel ("OJK Receives 153,000 Financial Scam Reports, 2025) menyebutkan bahwa Otoritas Jasa Keuangan menerima lebih dari 153.000 laporan penipuan keuangan, terutama terkait transaksi digital dan penyalahgunaan data pribadi, memperlihatkan meningkatnya kerentanan konsumen terhadap kejahatan finansial berbasis teknologi. Kondisi ini memperlihatkan bahwa sistem de-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengutipkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

teks penipuan harus mampu mengidentifikasi pola transaksi secara presisi pada lingkungan yang berubah sangat cepat.

Data transaksi keuangan memiliki karakteristik teknis yang menuntut metode analisis yang cermat. Setiap transaksi umumnya dipengaruhi oleh perilaku pengguna, jenis dan lokasi *merchant*, waktu transaksi, dan nilai pembayaran. Selain itu, ketidakseimbangan kelas (*class imbalance*) antara transaksi normal dan transaksi penipuan sangat signifikan, sehingga berpotensi menghasilkan model yang terlalu adaptif terhadap kelas mayoritas dan tidak cukup sensitif terhadap pola penipuan (Pozzolo, Boracchi, Caelen, Alippi, dan Bontempi, 2018) Studi oleh (Afriyie dkk., 2023) dan (Hafez, Hafez, Saleh, El-Mageed, dan Abohany, 2025) juga menegaskan bahwa metode konvensional berbasis aturan tidak lagi memadai untuk mengikuti perubahan pola *fraud* yang semakin adaptif, sehingga diperlukan pendekatan *machine learning* yang mampu menangani data berdimensi tinggi, distribusi kelas tidak seimbang, dan pola *non-linear*.

Berbagai algoritma *supervised learning* telah diterapkan dalam deteksi penipuan, termasuk *logistic regression* (Irawan, 2025), *support vector machine* (Gyamfi dan Abdulai, 2020), dan *decision tree* (Afriyie dkk., 2023), *gradient boosting tree* (Taha dan Malebary, 2020), dan *deep learning* (Raghavan dan Gayar, 2019). Di antara pendekatan tersebut, metode *Gradient Boosting Decision Tree* (GBDT) menunjukkan kinerja yang konsisten pada data berdimensi tinggi dan pola *non-linear* (Liu, Li, Gruyer, dan Tu, 2025). Tiga algoritma yang paling menonjol adalah XGBoost, LightGBM, dan CatBoost. XGBoost dikenal karena stabilitas prediksi dan kemampuan menangani dataset berukuran besar (Gupta dkk., 2023). LightGBM menggunakan strategi pertumbuhan pohon berbasis *leaf-wise* yang memberikan keunggulan dari sisi efisiensi komputasi (Xiao, Tan, dan Liu, 2025). CatBoost dirancang untuk menangani fitur kategorikal secara langsung tanpa melakukan *one-hot encoding*, sehingga menghasilkan struktur model yang stabil pada data transaksi keuangan (Babu, Maliakal, V, dan Babu, 2024). Dengan karakteristik tersebut, ketiga algoritma ini menjadi kandidat kuat untuk tugas deteksi penipuan pada data yang kompleks dan tidak seimbang.

Salah satu sumber data yang banyak dimanfaatkan untuk kajian deteksi penipuan kartu kredit adalah *Credit card Transactions Dataset* yang dikembangkan oleh (Shenoy, 2020) dengan lebih dari 1,8 juta baris data. Dataset ini merepresentasikan perilaku transaksi secara rinci, mencakup waktu transaksi, jenis *merchant*, karakteristik pengguna, dan nilai pembayaran, sehingga memungkinkan evaluasi kinerja model secara lebih realistis. Meskipun tidak secara spesifik merepresentasikan transaksi di Indonesia, struktur fiturnya yang kaya menjadikannya studi ka-

- per gul pre mo ma dip h b ters per tra 202 tive dita ber

per gul pre mo ma dip h b ters per tra 202 tive dita ber

per gul pre mo ma dip h b ters per tra 202 tive dita ber

per gul pre mo ma dip h b ters per tra 202 tive dita ber

per gul pre mo ma dip h b ters per tra 202 tive dita ber



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

retis yang kuat, bersifat *model-agnostic*, dan mampu memberikan interpretasi global maupun lokal secara konsisten, sehingga relevan digunakan untuk menganalisis karakteristik transaksi yang berpotensi sebagai penipuan (Noviandy, Idroes, Hardi, Afjal, dan Ray, 2024).

Berdasarkan uraian tersebut, masih terdapat kebutuhan akan penelitian yang secara sistematis membandingkan XGBoost, LightGBM, dan CatBoost pada dataset transaksi kartu kredit yang tidak seimbang, dengan mengintegrasikan SMOTE dan optimasi *hyperparameter* menggunakan GridSearchCV dalam suatu kerangka evaluasi yang terpadu, sekaligus menggabungkan analisis interpretabilitas berbasis SHAP untuk mengungkap karakteristik transaksi yang berpotensi sebagai penipuan. Penelitian ini diharapkan tidak hanya menghasilkan model dengan kinerja deteksi penipuan yang baik, tetapi juga memberikan pemahaman yang lebih mendalam mengenai kekuatan dan keterbatasan tiap algoritma, serta ciri-ciri utama yang membedakan transaksi penipuan dari transaksi normal, sehingga dapat menjadi dasar empiris yang relevan untuk pengembangan sistem deteksi penipuan pada industri keuangan.

1.2 Perumusan Masalah

Bagaimana perbandingan performa algoritma CatBoost, LightGBM, dan XGBoost dalam mendeteksi penipuan kartu kredit pada data yang tidak seimbang, dengan mempertimbangkan efektivitas teknik *oversampling* SMOTE, *hyperparameter tuning*, dan analisis interpretabilitas berbasis SHAP, untuk mengetahui algoritma mana yang memberikan hasil paling optimal dan stabil?

1.3 Batasan Masalah

Batasan masalah penelitian ini adalah:

1. Penelitian ini hanya menggunakan dataset *public* transaksi kartu kredit yang dari Kaggle, yaitu <https://www.kaggle.com/datasets/kartik2112/fraud-detection>. Dataset ini terdiri dari 1.852.394 baris data, dari transaksi kartu kredit yang disimulasikan antara 1 Januari 2019 hingga 31 Desember 2020, yang mencakup transaksi sah maupun transaksi penipuan yang berlokasi di wilayah bagian barat Amerika Serikat.
2. Penelitian ini tidak akan melibatkan implementasi langsung dari hasil *machine learning* ke dalam praktis klinis, melainkan memberikan bukti konseptual tentang model yang digunakan.
3. Pengujian dilakukan dengan menggunakan algoritma CatBoost, XGBoost, dan LightGBM dengan metode *oversampling* SMOTE dan *hyperparameter tuning* GridSearchCV.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. Fitur yang digunakan dalam pemodelan dibatasi pada:
 - a. Atribut asli (raw): 'amt', 'category', 'gender', dan 'state'.
 - b. Atribut hasil *feature engineering*: 'trans_hour', 'trans_dayofweek', 'day_period', 'is_weekend', 'age', 'distance_km', 'diff_amt', 'avg_amt_last_7', 'count_last_7', 'category_freq', dan 'merchant_risk_rate'.
 - c. Atribut lain yang bersifat identitas unik atau berpotensi menimbulkan data leakage (misalnya 'cc_num', 'first', 'last', 'street', 'dob', 'trans_num', dan 'unix_time') tidak digunakan sebagai fitur pemodelan.
5. Kinerja algoritma dievaluasi menggunakan metrik performa yang dihitung dari *confusion matrix*, meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Serta metrik *Area Under the Receiver Operating Characteristic curve* (AUC-ROC) untuk menilai kemampuan model dalam membedakan kelas.
6. Analisis interpretabilitas dalam penelitian ini dibatasi pada penggunaan *Shapley Additive Explanations* (SHAP) sebagai pendekatan XAI untuk menjelaskan kontribusi fitur terhadap prediksi model.

1.4 Tujuan

Tujuan penelitian ini adalah:

1. Tujuan penelitian ini adalah menganalisis dan membandingkan kinerja tiga algoritma *boosting modern* XGBoost, LightGBM, dan CatBoost dalam mendeteksi penipuan kartu kredit pada dataset yang sangat tidak seimbang, dengan menguji pengaruh penerapan teknik *oversampling* SMOTE serta optimasi *hyperparameter* terhadap metrik evaluasi utama, seperti *Accuracy*, *Precision*, *Recall*, *F1-score*, dan AUC-ROC.
2. Mengkaji interpretabilitas model melalui pendekatan *Shapley Additive Explanations* (SHAP) guna mengidentifikasi fitur-fitur utama yang memengaruhi klasifikasi transaksi sebagai *fraud*, serta memberikan transparansi terhadap proses prediksi model.
3. Memberikan rekomendasi mengenai kombinasi model dan teknik prapemrosesan yang paling efektif dan efisien untuk studi kasus deteksi penipuan.

1.5 Manfaat

Manfaat penelitian ini adalah:

1. Dengan membandingkan kedua algoritma dalam menangani data yang tidak seimbang, penelitian ini akan memberikan wawasan terkait teknik *oversampling* data dan *hyperparameter tuning* yang efektif. Dan juga dapat membuka penelitian terbaru kedepannya guna menentukan akurasi model yang



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

lebih baik dari sebelumnya.

2. Menyajikan wawasan berbasis bukti mengenai keunggulan dan kelemahan setiap model, yang dapat menjadi landasan pertimbangan bagi praktisi dalam merancang dan memilih strategi deteksi penipuan yang lebih efektif.

1.6 Sistematika Penulisan

Penulisan dalam penelitian ini disusun dengan sistematika sebagai berikut:

BAB 1. PENDAHULUAN

Pada bagian ini memberikan gambaran mengenai latar belakang, perumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, dan struktur penulisan laporan tugas akhir.

BAB 2. LANDASAN TEORI

Pada bagian ini menjelaskan mengenai tentang teori-teori yang menjadi dasar dalam tugas akhir.

BAB 3. METODOLOGI PENELITIAN

Bagian ini mencakup pelaksanaan dari penelitian tugas akhir yang mencakup perencanaan, pengumpulan data, alur penelitian, pengelolaan data, dan analisis data.

BAB 4. ANALISA DAN HASIL

Pada bagian ini menyajikan hasil penelitian yang telah dilakukan serta pembahasannya secara rinci dari proses pengumpulan data, pengolahan data dan analisis data.

BAB 5. PENUTUP

Bagian ini berisi kesimpulan dari hasil penelitian yang telah diperoleh, serta saran yang dapat dijadikan acuan bagi penelitian selanjutnya maupun pengembangan sistem deteksi penipuan pada konteks praktis.

BAB 2

LANDASAN TEORI

2.1 Machine Learning

Machine Learning adalah cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan algoritma dan model statistik yang memungkinkan sistem untuk belajar dari data dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit (Okagbue dkk., 2023). Klasifikasi dalam *machine learning* merupakan proses penting yang memungkinkan sistem untuk mengkategorikan data berdasarkan pola yang telah dipelajari dari data sebelumnya. Terdapat empat pendekatan utama dalam klasifikasi *machine learning*, yaitu *supervised learning*, *unsupervised learning*, *semi-supervised learning*, dan *reinforcement learning* (Matsuo dkk., 2020).

2.1.1 Supervised Learning

Supervised Learning melibatkan pelatihan model menggunakan dataset yang telah diberi label. Model belajar memetakan input ke output berdasarkan contoh-contoh yang diberikan. Pendekatan ini efektif untuk tugas-tugas seperti klasifikasi dan regresi. Contoh algoritma yang umum digunakan meliputi Support Vector Machines (SVM), Decision Trees, dan Neural Networks. Keunggulan utama dari *supervised learning* adalah kemampuannya untuk menghasilkan prediksi yang akurat ketika data pelatihan representatif dan berkualitas tinggi (Matsuo dkk., 2020).

2.1.2 Unsupervised Learning

Unsupervised learning digunakan ketika data tidak memiliki label. Model berusaha menemukan struktur atau pola tersembunyi dalam data. Teknik ini sering digunakan untuk clustering dan reduksi dimensi. Algoritma seperti K-Means dan *Principal Component Analysis* (PCA) adalah contoh dari pendekatan ini. Meskipun tidak memerlukan data berlabel, tantangan utama dalam *unsupervised learning* adalah interpretasi hasil yang mungkin tidak selalu jelas (Mathivanan, Md.Ghani, dan Janor, 2019).

2.1.3 Semi-Supervised Learning

Semi-supervised learning merupakan kombinasi antara *supervised* dan *unsupervised learning*. Model dilatih menggunakan sejumlah kecil data berlabel dan sejumlah besar data tanpa label. Pendekatan ini berguna ketika pelabelan data mahal atau memakan waktu. Dengan memanfaatkan informasi dari data tanpa label, mod-

el dapat meningkatkan akurasi prediksi dibandingkan dengan menggunakan data berlabel saja (Ramírez-Sanz, Maestro-Prieto, Álar Arnaiz-González, dan Bustillo, 2023).

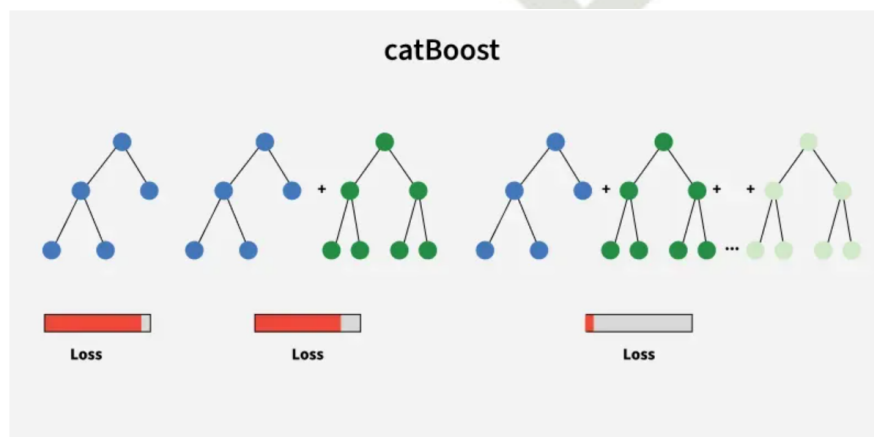
2.1.4 Reinforcement Learning

Reinforcement learning melibatkan agen yang belajar melalui interaksi dengan lingkungan. Agen menerima umpan balik dalam bentuk *reward* atau *punishment* berdasarkan tindakan yang diambil. Tujuan utama adalah mempelajari kebijakan tindakan yang memaksimalkan *reward* kumulatif. Pendekatan ini banyak digunakan dalam pengembangan sistem otonom, seperti robotika dan permainan computer (Vincent, Peter, Riashat, Marc, dan Joelle, 2018).

2.2 Categorical Boosting (CatBoost)

Berakar pada keluarga *Gradient Boosted Decision Trees* (GBDT), CatBoost dengan cepat menjadi salah satu algoritma pilihan utama untuk tugas klasifikasi terawasi (Dev dan Eden, 2019). CatBoost adalah algoritma *gradient boosting* yang dioptimalkan untuk menangani fitur kategorikal menggunakan ordered boosting guna mencegah *overfitting* (Y. Chen dan Han, 2021). Pada Persamaan 2.1, fungsi prediksi keseluruhan $F(x)$ pada metode CatBoost dibentuk dengan menggabungkan tebakan awal $F^0(x)$, yang umumnya merupakan nilai rata-rata dari variabel target, dengan kontribusi setiap pohon keputusan $F_m(x_i)$ untuk masing-masing sampel pelatihan. Seluruh kontribusi tersebut kemudian dijumlahkan dalam suatu *ensemble* yang terdiri atas M pohon dan N sampel pelatihan (Babu dkk., 2024). Struktur CatBoost dapat dilihat pada Gambar 2.1.

$$F(x) = F^0(x) + \sum_{m=1}^M \sum_{i=1}^N F_m(x_i) \quad (2.1)$$



Gambar 2.1. Struktur CatBoost



Hak Cipta Dilindungi Undang-Undang

CatBoost didasarkan pada konsep teknik *gradient boosting*, di mana pohon keputusan dibangun secara berurutan untuk meminimalkan *error* dan meningkatkan akurasi prediksi. Prosesnya dimulai dengan membangun sebuah pohon keputusan dan mengevaluasi besarnya kesalahan prediksi. Setelah pohon pertama dibuat, pohon berikutnya dibangun untuk mengoreksi kesalahan yang dihasilkan oleh pohon sebelumnya. Tahap ini berlangsung secara iteratif, dengan setiap pohon baru berfokus untuk memperbaiki prediksi model melalui pengurangan residual dari iterasi sebelumnya, hingga jumlah iterasi yang telah ditentukan tercapai. Hasil akhirnya adalah sebuah *ensemble* pohon keputusan yang bekerja bersama-sama untuk menghasilkan prediksi yang lebih akurat (Detthamrong, Chansanam, Boongoen, dan Iam-On, 2024) (GeeksforGeeks, 2025).

CatBoost sangat cocok digunakan pada dataset berskala besar yang memiliki banyak fitur independen. Berbeda dari algoritma *gradient boosting* lainnya, CatBoost dirancang secara khusus untuk menangani fitur kategorikal dan numerik secara langsung tanpa memerlukan proses encoding manual (M., K., R., AboRas, dan Youssef, 2024).

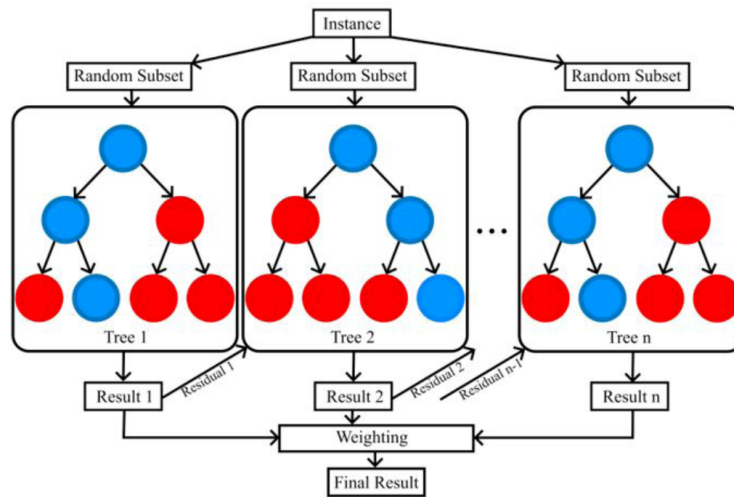
2.3 Extreme Gradient Boosting (XGBoost)

XGBoost adalah algoritma yang sangat efisien dan efektif untuk mengatasi masalah dalam skala besar. Algoritma ini menggunakan pendekatan penguatan pohon *gradien*, yang menggabungkan beberapa pohon keputusan untuk secara bertahap mengoreksi kesalahan pohon sebelumnya hingga akurasi optimal tercapai (Sankar, Potti, Chandrika, dan Ramasubbareddy, 2022). XGBoost dikenal memiliki performa unggul dalam meningkatkan akurasi, efisiensi waktu pelatihan, serta pemanfaatan sumber daya komputasi. Selain itu, algoritma ini juga efektif dalam meminimalkan risiko *overfitting* (Scarano dkk., 2025).

Struktur kerja XGBoost ditunjukkan pada Gambar 2.2, di mana model dibangun secara bertahap menggunakan subset data acak (*random subset*). Pohon pertama menghasilkan prediksi awal, lalu residu dari hasil tersebut digunakan untuk melatih pohon berikutnya, dan proses ini berlanjut hingga terbentuk sejumlah pohon keputusan. Prediksi akhir tidak diperoleh melalui *majority vote*, melainkan dari kombinasi berbobot (*weighting*) setiap pohon sesuai tingkat kepentingannya (Oztornaci, Ata, dan Kartal, 2024).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2.2. Struktur XGBoost

Perhitungan prediksi dalam XGBoost dijelaskan melalui suatu persamaan tertentu (Karo, 2020). Bentuk persamaan algoritma XGBoost dapat dilihat pada Persamaan (2.2), Persamaan (2.3), dan Persamaan (2.4).

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \quad (2.2)$$

$$\mathcal{F} = \{f(x) = w_{q(x)}\}, \quad q: \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T \quad (2.3)$$

$$L(\theta) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2.4)$$

Keterangan:

$\hat{y}_i$: Nilai prediksi dari model untuk sampel ke- i .

f_k : Pohon keputusan ke- k dari total K pohon dalam *ensemble*.

$\mathcal{F}$: Ruang fungsi dari semua pohon regresi.

$f(x) = w_{q(x)}$: Fungsi yang menghasilkan nilai prediksi dari bobot daun w berdasarkan struktur pohon $q(x)$.

$q: \mathbb{R}^m \rightarrow T$: Fungsi pemetaan dari fitur input ke indeks daun.

$w \in \mathbb{R}^T$: Bobot pada setiap daun dari pohon.

T : Jumlah daun pada satu pohon.

$l(\hat{y}_i, y_i)$: Fungsi *loss* antara nilai prediksi dan nilai aktual.

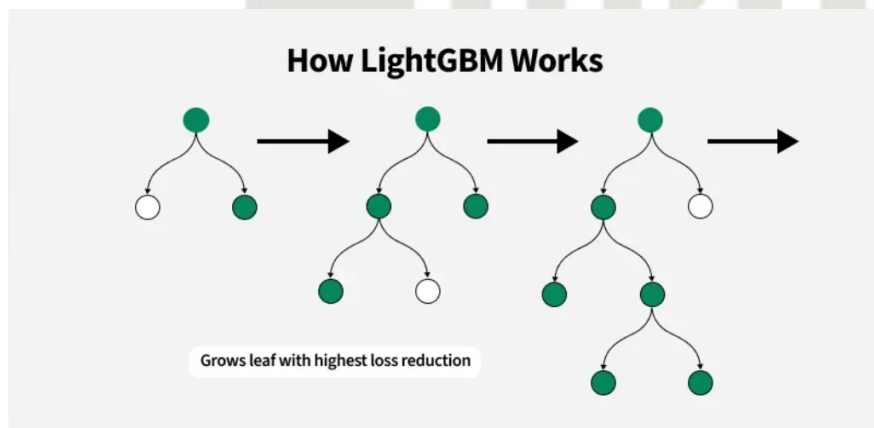
$\Omega(f_k)$: Fungsi kompleksitas dari pohon ke- k (biasanya berupa regularisasi).

2.4 Light Gradient Boosting Machine (LightGBM)

Light Gradient Boosting Machine merupakan algoritma *gradient boosting* berbasis pohon keputusan yang dikembangkan untuk mengatasi keterbatasan GB-DT konvensional yang cenderung lambat dan membutuhkan sumber daya komputasi besar ketika berhadapan dengan dataset berukuran besar. Pada GB-DT tradisional, proses pencarian titik pemisahan (*split point*) terbaik dilakukan dengan mengevaluasi seluruh sampel dan seluruh fitur, sehingga membuat proses pelatihan menjadi cukup intensif secara komputasi (Detthamrong dkk., 2024).

Untuk meningkatkan efisiensi proses tersebut, LightGBM memperkenalkan dua teknik utama, yaitu *Gradient-based One-Side Sampling* (GOSS) dan *Exclusive Feature Bundling* (EFB). GOSS memilih subset sampel berdasarkan nilai gradien sehingga hanya sampel yang berkontribusi besar terhadap fungsi *loss* yang diprioritaskan. Sementara itu, EFB menggabungkan fitur-fitur yang saling eksklusif ke dalam satu bundel sehingga mengurangi dimensi fitur tanpa kehilangan informasi penting.

Selain optimasi tersebut, LightGBM menggunakan strategi pertumbuhan berbasis daun (*leaf-wise*), di mana node dengan peningkatan (*gain*) tertinggi dipilih terlebih dahulu untuk dikembangkan. Berbeda dengan metode *level-wise* pada algoritma pohon lain, strategi ini memungkinkan model mencapai akurasi lebih tinggi dengan jumlah iterasi yang lebih sedikit, meskipun berpotensi menghasilkan struktur pohon yang lebih dalam. LightGBM juga melakukan diskretisasi fitur kontinu ke dalam sejumlah bin sehingga proses pemilihan split pada setiap node menjadi lebih efisien. Struktur LightGBM dapat dilihat pada Gambar 2.3.



Gambar 2.3. Struktur LightGBM

Gain atau perolehan informasi dari suatu pemisahan dihitung menggunakan fungsi nilai seperti ditunjukkan pada Persamaan (5). Pada formula tersebut, (O)

Hak Cipta Dilindungi Undang-Undang

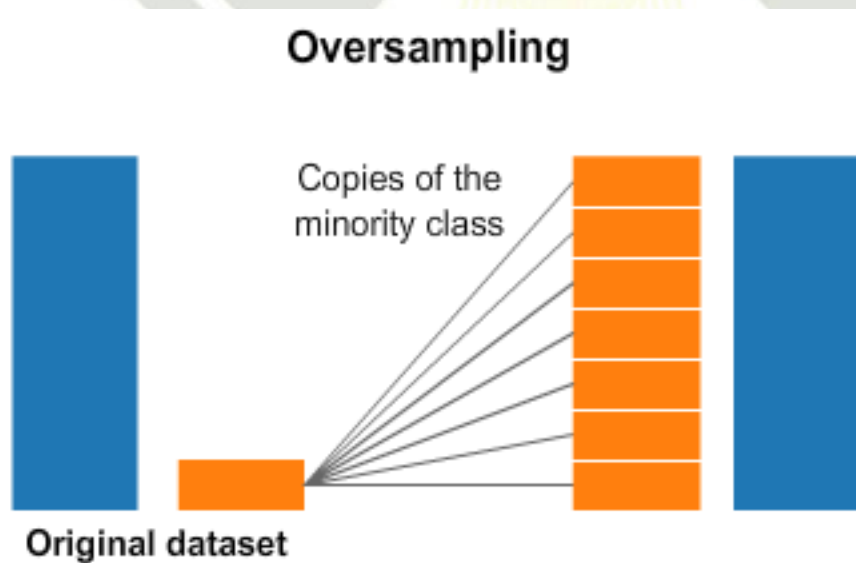
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

merepresentasikan himpunan sampel pada node tertentu, (j) merupakan indeks fitur yang diuji sebagai kandidat pemisah, dan (d) adalah nilai ambang (*threshold*) untuk memisahkan sampel menjadi node kiri dan kanan. Split dipilih ketika nilai gain yang dihitung merupakan yang terbesar di antara seluruh kandidat, sehingga node akan dibagi menjadi node anak kiri (*left child*) dan node anak kanan (*right child*) (Noh dan Kang, 2025).

$$V_{j|o}(d) = \frac{1}{n_o} \left(\frac{\left(\sum_{\{x_i \in O: x_{ij} \leq d\}} g_i \right)^2}{n_{l|o}^j} + \frac{\left(\sum_{\{x_i \in O: x_{ij} > d\}} g_i \right)^2}{n_{r|o}^j} \right) \quad (2.5)$$

2.5 Synthetic Minority Over-Sampling Technique (SMOTE)

Metode *oversampling* yang disebut SMOTE digunakan untuk mengatasi ketidakseimbangan data. Dengan memanfaatkan tetangga terdekatnya berdasarkan jarak *Euclidean* untuk membuat data sintetis dari kelas minoritas. SMOTE meningkatkan jumlah sampel karena data sintetis dibuat menggunakan fitur yang sudah ada sebelumnya, sehingga terlihat mirip dengan data asli (R. Chen dkk., 2023). Visualisasi dari *Oversampling* ini dapat dilihat pada Gambar 2.4.



Gambar 2.4. Visualisasi *Oversampling*

Bentuk persamaan SMOTE dapat dilihat pada Peramaan (2.6).

$$x_{\text{new}} = x_i + \lambda(x_j - x_i) \quad (2.6)$$



Keterangan:

x_{new} : Data sintetis (baru) yang dihasilkan.

x_i : Sampel ke- i yang dipilih dari kelas minoritas.

x_j : Sampel ke- j yang dipilih secara acak dari kelas minoritas.

λ : Bilangan acak yang diambil dari distribusi *uniform*.

2.6 Hyperparameter Tuning dengan Grid Search Cross Validation

Hyperparameter tuning adalah nilai-nilai yang ditetapkan sebelum model dilatih dan tidak dipelajari langsung dari data, melainkan diatur sebelum proses pelatihan untuk menentukan struktur serta perilaku model. Untuk memperoleh prediksi yang akurat dan model yang efisien, tidak cukup hanya memilih algoritma, pengaturan *hyperparameter* yang optimal melalui proses *tuning* menjadi kritis (Fadzail, Zali, Mid, dan Bakar, 2025) (Raiaan dkk., 2024). Proses *hyperparameter tuning* bertujuan mencari kombinasi nilai terbaik agar performa model dalam hal akurasi, kemampuan generalisasi, atau efektivitas dapat dimaksimalkan. (Ogunsanya, Isichei, dan Desai, 2023) menekankan bahwa *tuning* adalah "the process of selecting the combination of optimal hyperparameters that give the best performance to the model" artinya proses pemilihan kombinasi *hyperparameter* yang optimal untuk memberikan performa terbaik pada model.

Grid Search Cross-Validation (GridSearchCV) merupakan salah satu metode yang digunakan untuk menentukan kombinasi parameter terbaik dalam suatu model. Metode ini bekerja dengan mengevaluasi setiap kemungkinan kombinasi parameter secara sistematis, lalu melakukan validasi terhadap masing-masing kombinasi tersebut. GridSearchCV akan memberikan hasil yang optimal jika rentang nilai minimum dan maksimum dari setiap parameter telah ditentukan sebelumnya (RAMADHAN, SITANGGANG, NASUTION, dan GHIFARI, 2021) (Verwaeren, der Weeën, dan Baets, 2022).

2.7 Explainable Artificial Intelligent (XAI)

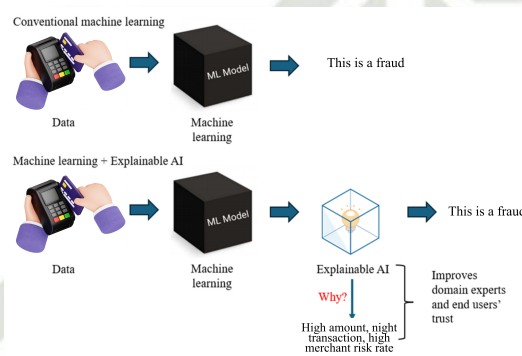
Explainable Artificial Intelligence (XAI) semakin mendapat perhatian sebagai pendekatan untuk mengatasi permasalahan interpretabilitas pada model *machine learning* (Ali dkk., 2023) (Saeed dan Omlin, 2023). Ketika algoritma yang digunakan semakin kompleks dan canggih, upaya untuk menjadikannya dapat dijelaskan (*explainable*) menjadi tantangan tersendiri. Untuk mengatasi permasalahan ini, dikembangkanlah metode *Explainable Artificial Intelligence* (XAI). XAI merupakan pendekatan tambahan pada model *machine learning* yang tidak memengaruhi

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

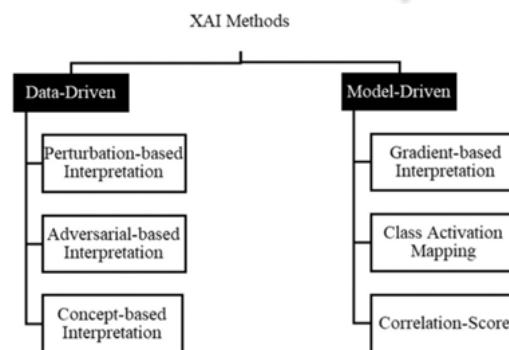
akurasi maupun proses komputasi dari model tersebut. Sebaliknya, XAI bertujuan untuk mendukung proses pengambilan keputusan dengan memberikan penjelasan atas alasan di balik prediksi yang dihasilkan.

XAI mengembangkan seperangkat alat, teknik, dan strategi yang menghasilkan model yang lebih transparan, akuntabel, dan dapat dijelaskan, sekaligus tetap mempertahankan kemampuan prediksi yang kuat dan memberikan kepercayaan terhadap keluaran model (Roshan dan Zafar, 2021). Gambar 2.5 menggambarkan bagaimana integrasi XAI mampu meningkatkan metode *machine learning* konvensional dengan mengungkapkan logika atau proses berpikir di balik prediksi model.



Gambar 2.5. Penggunaan XAI di ML

XAI memungkinkan model yang semula bersifat *black-box* (tertutup dan tidak transparan) menjadi lebih seperti *glass-box* (terbuka dan dapat dijelaskan), sehingga alasan di balik prediksi dapat dipahami dengan jelas hal ini sangat penting dalam berbagai proses pengambilan keputusan yang kritis (Ukwaththa, Herath, dan Meddage, 2024). Seperti yang ditampilkan dalam Gambar 2.6, metode XAI umumnya diklasifikasikan ke dalam dua pendekatan utama, yaitu berbasis data (*data-driven*) dan berbasis model (*model-driven*).



Gambar 2.6. Klasifikasi metode Explainable AI (XAI)

Metode XAI berbasis data bekerja dengan memodifikasi data input dan mengamati dampaknya terhadap prediksi model. Pendekatan ini terdiri dari tiga jenis (Liang, Li, Yan, Li, dan Jiang, 2021), yaitu:

1. *Perturbation-based*: Mengubah atau menutupi sebagian data, kemudian menganalisis dampaknya terhadap hasil prediksi model.
2. *Adversarial-based*: Menambahkan data gangguan (*adversarial*) ke dalam dataset untuk menguji sensitivitas model terhadap perubahan yang halus.
3. *Concept-based*: Melibatkan konsep atau domain *knowledge* tertentu sebagai acuan interpretasi, biasanya melibatkan pengetahuan dari para ahli.

Di sisi lain, pendekatan XAI berbasis model lebih menitikberatkan pada mekanisme internal serta proses pengambilan keputusan dari model itu sendiri. Pendekatan ini bertujuan untuk memahami bagaimana suatu prediksi dihasilkan dengan mengamati struktur atau logika di dalam model. Salah satu metode yang umum digunakan adalah analisis *feature importance*, yaitu mengukur kontribusi relatif masing-masing fitur terhadap hasil prediksi model, sehingga dapat diketahui fitur mana yang memiliki pengaruh dominan dalam keputusan model. Selain itu, teknik visualisasi model juga sering digunakan, seperti *decision tree* atau *saliency map*, yang mampu menyajikan representasi visual secara intuitif mengenai batas-batas keputusan serta area input yang paling memengaruhi hasil keluaran model (A. dan R., 2023) (Ukwaththa dkk., 2024).

Menurut (Liang dkk., 2021) metode XAI berbasis model ini umumnya dikelompokkan ke dalam tiga pendekatan utama, yaitu pendekatan berbasis *gradient*, *class activation mapping*, dan analisis skor korelasi. Ketiga pendekatan tersebut menawarkan cara yang berbeda untuk menelusuri proses internal model dan menjelaskan dasar prediksi yang dihasilkan, sehingga dapat memberikan transparansi dan kepercayaan dalam implementasi sistem berbasis *machine learning*.

2.8 Shapley Additive Explanations (SHAP)

Shapley Additive exPlanations (SHAP) merupakan pendekatan terpadu untuk interpretasi model yang diperkenalkan oleh (Lundberg dan Lee, 2020). Metode ini bersifat *model-agnostic* dan dikembangkan untuk menjawab tantangan interpretabilitas pada model-model kompleks, seperti *ensemble methods* dan *deep neural networks*. Kerangka kerja SHAP mengintegrasikan beberapa teknik yang telah diajarkan sebelumnya, antara lain LIME dan DeepLIFT, ke dalam kelas *additive feature attribution*, yaitu keluarga metode di mana model penjelas direpresentasikan sebagai fungsi linear dari variabel-variabel biner (Roshan dan Zafar, 2021).

Secara konseptual, SHAP bertujuan menjelaskan prediksi model untuk su-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

atu input tertentu dengan menghitung kontribusi masing-masing fitur terhadap prediksi tersebut. Untuk mencapai tujuan ini, SHAP memanfaatkan *Shapley Values*, sebuah konsep yang berasal dari *coalitional game theory* (Bifarin, 2023). Dalam teori permainan, *Shapley Value* merupakan metode untuk mendistribusikan “imbalan” (*reward*) kepada para pemain berdasarkan kontribusi mereka terhadap total keuntungan. Ide ini kemudian diadaptasi dalam SHAP untuk menilai fitur mana yang berkontribusi signifikan terhadap keluaran model. Secara formal, *Shapley Value* didefinisikan sebagai rata-rata kontribusi marginal suatu fitur terhadap prediksi, dihitung pada seluruh kemungkinan koalisi fitur (Guo, Yang, Guo, dan Shen, 2024) (Sun, Kraus, Eilts, dan Garus, 2025). Secara khusus, pentingnya suatu fitur i dinyatakan melalui nilai Shapley sesuai Persamaan (2.7) (Molnar, 2020).

$$\phi_i = \sum_{S \subseteq Z \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)] \quad (2.7)$$

Pada persamaan tersebut, S merepresentasikan semua subset fitur yang tidak mengandung fitur i , M adalah jumlah total fitur, dan koefisien pembobot $\frac{|S|!(M - |S| - 1)!}{M!}$ memastikan pembobotan yang adil terhadap kombinasi fitur dari berbagai ukuran. Rumus ini menghitung dampak rata-rata fitur i terhadap prediksi model dengan menghitung kontribusi marjinalnya di seluruh kemungkinan kombinasi fitur, kemudian mengambil nilai rata-rata berbobot. Perumusan ini juga mempertimbangkan urutan fitur, yang menjadi penting ketika fitur memiliki korelasi, karena urutan fitur dapat memengaruhi kontribusi yang dihitung.

Berdasarkan sifat *efficiency* (Molnar, 2020), metode SHAP juga menjamin bahwa setiap prediksi individu dapat diuraikan sepenuhnya menjadi kontribusi fitur-fitur penyusunnya, sebagaimana ditunjukkan pada Persamaan (2.8).

$$f(i) = \phi_0 + \sum_j \phi_{i,j} \quad (2.8)$$

Di mana $f(i)$ adalah nilai prediksi untuk sampel ke- i , ϕ_0 menyatakan nilai dasar (*baseline*), dan $\phi_{i,j}$ merupakan nilai SHAP dari fitur ke- j terhadap prediksi untuk sampel ke- i . Persamaan ini memastikan bahwa total kontribusi seluruh fitur terhadap satu prediksi setara dengan selisih antara nilai prediksi dan nilai dasar, sehingga memungkinkan kuantifikasi yang tepat terhadap pengaruh masing-masing fitur terhadap prediksi individu.

Dalam penelitian ini, analisis *SHAP* digunakan untuk menjelaskan model pada dua tingkat, yaitu tingkat global dan tingkat individu. Pada tingkat global, grafik batang *feature importance* digunakan untuk menunjukkan peringkat fi-

tur berdasarkan nilai rata-rata absolut dari $|\phi_{i,j}|$, yang menggambarkan pengaruh masing-masing fitur terhadap model secara keseluruhan.

2.9 Confusion Matrix

Confusion matrix adalah suatu metode evaluasi yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap label aktual. *Confusion matrix* memberikan informasi mengenai bagaimana suatu input diklasifikasikan ke dalam kelas yang berbeda. Tabel 2.1 menampilkan *confusion matrix* (Faraji, 2022).

Tabel 2.1. Confusion Matrix

<i>Confusion Matrix</i>		
	<i>Positive (Fraud)</i>	<i>Negative (Normal)</i>
<i>Positive (Fraud)</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Negative (Normal)</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Metrik kinerja dapat dihitung berdasarkan *confusion matrix* sebagai berikut:

1. *True Negative (TN)* = jumlah transaksi sah yang berhasil diprediksi sebagai transaksi sah.
2. *False Positive (FP)* = jumlah transaksi sah yang salah diprediksi sebagai penipuan.
3. *False Negative (FN)* = jumlah transaksi penipuan yang salah diprediksi sebagai transaksi sah.

Melalui *confusion matrix*, dapat dilihat seberapa baik model dalam membedakan antara transaksi yang benar-benar penipuan dengan transaksi yang sah. Evaluasi ini kemudian diperkuat dengan penggunaan metrik penilaian seperti *Accuracy*, *Precision*, *Recall*, dan *F1-score* yang masing-masing dirumuskan pada Persamaan (2.9), Persamaan (2.10), Persamaan (2.11), dan Persamaan (2.12).

Metrik-metrik tersebut memberikan gambaran menyeluruh mengenai kelebihan dan kelemahan model, sehingga dapat memastikan bahwa algoritma yang dipilih merupakan solusi yang paling optimal. Selain itu, hasil evaluasi juga mencerminkan efektivitas proses optimasi model, termasuk penyesuaian parameter algoritma yang dilakukan untuk meningkatkan performa (Shobayo, Zachariah, Odusami, dan Ogunleye, 2023) (Chang, Ali, Golightly, Ganatra, dan Mohamed, 2024).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.9)$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2.10)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.11)$$

$$F1-Score = \frac{2TP}{2TP + FP + FN} \times 100\% \quad (2.12)$$

2.10 Area Under the Curve

Area under the curve (AUC) adalah metrik krusial yang berasal dari kurva ROC untuk mengukur kemampuan suatu model klasifikasi dalam melakukan perbedaan antara kelas-kelas yang berbeda. Nilai AUC yang tinggi mengindikasikan bahwa model tersebut mempunyai performa baik dalam memisahkan *instance positif* dan *instance negatif* dengan tingkat akurasi yang lebih tinggi. Nilai ini berkisar dari 0 sampai 1, dengan nilai 0 menandakan model yang salah mengklasifikasikan semua data, sementara nilai 1 menandakan model ideal yang berhasil mengklasifikasikan semua data secara tepat (Z. Zhang, 2016). Semakin dekat nilainya ke angka 1, maka model tersebut semakin memperlihatkan kapabilitas yang kuat dalam membedakan kelas dengan akurat (Andrade-Arenas dan Yactayo-Arias, 2025).

Semakin besar nilai AUC-ROC, semakin besar kemungkinan bahwa sampel positif akan ditempatkan di depan sampel negatif. Persamaan (2.13) berikut diberikan sebagai salah satu rumus perhitungannya.

$$AUC-ROC = \frac{\sum \text{rank}_{n \in \text{positive}} - \frac{P(1+P)}{2}}{PQ} \quad (2.13)$$

di mana P merepresentasikan jumlah sampel positif, Q menyatakan jumlah sampel negatif, dan rank_n merupakan hasil peringkat dari sampel positif ke- n berdasarkan probabilitas prediksi.

2.11 Deteksi Penipuan Kartu Kredit

Penipuan kartu kredit (*credit card fraud*) merupakan penggunaan kartu kredit secara tidak sah dengan tujuan memperoleh sesuatu yang bernilai disertai maksud untuk menipu. Definisi ini sejalan dengan rumusan *National White Collar Crime* (Prasetyo, 2021) menekankan unsur penggunaan tanpa otorisasi serta adanya *intent to defraud*. Lebih lanjut, (Bhatla, 2003) menjelaskan bahwa fraud terjadi ketika seseorang menggunakan kartu kredit milik pihak lain untuk kepentingan pribadi sementara pemilik kartu dan penerbit tidak menyadari bahwa kartu tersebut sedang digunakan. Pelaku juga tidak memiliki hubungan dengan pemegang kartu



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

maupun penerbit, serta tidak memiliki itikad baik untuk menghubungi pemilik atau melakukan pembayaran atas transaksi yang dilakukan (Bhatla, 2003).

Karena pelaku berupaya membuat transaksi terlihat wajar, maka suatu transaksi umumnya tidak dinyatakan sebagai penipuan hanya dari satu ciri tunggal, melainkan dari kombinasi indikasi atau indikator yang menyimpang dari kebiasaan pengguna dan konteks transaksi normal (Adepoju dan Wosowei, 2019). Secara proses tindak *credit fraud* sering diawali oleh *identity theft*, yaitu penguasaan data identitas kartu secara ilegal seperti nama penerbit kartu, nama pemegang kartu, nomor kartu, masa berlaku kartu, maupun nomor verifikasi kartu atau CVV (*Card Verification Value*), yang kemudian digunakan untuk melakukan transaksi tanpa izin pemegang sah (Adepoju dan Wosowei, 2019). Implikasinya, indikator *fraud* lebih tepat dipahami sebagai pola transaksi yang menunjukkan adanya penyalahgunaan otorisasi, baik pada transaksi berbasis kartu fisik (*card-present*) maupun transaksi jarak jauh atau *online* (*card-not-present*) yang hanya memerlukan informasi penting kartu (Prasetyo, 2021).

Dalam konteks penelitian, deteksi *fraud* banyak dijadikan studi empiris karena karakteristik datanya yang khas, terutama ketidakseimbangan kelas (*class imbalance*) ketika transaksi normal jauh lebih dominan daripada transaksi penipuan, serta pola *fraud* yang adaptif dan dapat berubah dari waktu ke waktu (Majumder, 2024) (Pozzolo dkk., 2018). Kondisi ini menjadikan *fraud* detection relevan untuk menguji kinerja algoritma *machine learning*, termasuk metode *ensemble* dan *boosting*, dalam kerangka evaluasi yang konsisten (Gupta dkk., 2023).

Secara operasional, dasar atau indikator transaksi beresiko *fraud* dapat diturunkan ke beberapa dimensi pola yang lazim dipantau dalam sistem pembayaran, sebagai berikut:

1. Indikator berbasis akses/kanal transaksi (*card-present vs card-not-present*)

Fraud dapat muncul pada transaksi yang melibatkan kartu fisik maupun pada transaksi jarak jauh yang hanya mengandalkan data kartu. Perbedaan kanal ini penting karena transaksi jarak jauh cenderung bergantung pada kerahasiaan data (nomor kartu, masa berlaku, CVV), sehingga pola penyalahgunaan sering beririsan dengan kebocoran/penguasaan data (Apriwandi dan Herycson, 2022) (Fardiansyah, Satory, dan Dewi, 2022) (Oluwatoyin dan Akinola, 2024).

2. Indikator berbasis penguasaan data (*identity theft sebagai prasyarat*)

Transaksi berindikasi *fraud* umumnya terkait dengan adanya penguasaan kartu/data kartu secara melawan hukum, baik melalui pencurian kartu fisik



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

maupun perolehan data elektronik (misalnya melalui skimming dan teknik lain). Kerangka ini menegaskan bahwa indikator *fraud* di transaksi sering merupakan “kelanjutan” dari aktivitas *identity theft* (Fardiansyah dkk., 2022) (Prasetyo, 2021).

3. **Indikator berbasis waktu transaksi**

Literatur modus operandi menunjukkan kecenderungan pelaku memanfaatkan waktu tertentu ketika pengawasan dan verifikasi relatif longgar, sehingga transaksi pada jam yang tidak lazim atau pada rentang waktu tertentu dapat menjadi sinyal pendukung, terutama bila konsisten berulang dan muncul bersama indikator lain (idris, 2020).

4. **Indikator berbasis nilai/nominal transaksi**

Pola *fraud* dapat tercermin pada nilai transaksi yang menyimpang misalnya lonjakan nilai yang tidak proporsional atau adanya perubahan nilai pada dokumen transaksi. Karena *fraud* bertujuan memperoleh keuntungan finansial, dimensi nilai transaksi menjadi salah satu indikator yang relevan untuk ditinjau secara pola (Prasetyo, 2021).

5. **Indikator berbasis intensitas dan pemecahan transaksi (*split charge*)**

Salah satu pola yang dicatat dalam pedoman penindakan adalah upaya menghindari otorisasi/ambang verifikasi melalui pemecahan transaksi menjadi beberapa transaksi dengan nominal tertentu. Secara analitis, pola ini muncul sebagai frekuensi transaksi yang meningkat dalam interval waktu singkat dan komposisi nominal yang “terfragmentasi” (Prasetyo, 2021).

6. **Indikator berbasis merchant dan dokumen/proses transaksi**

Fraud dapat terkait dengan konteks merchant atau penyalahgunaan proses, misalnya penggandaan sales draft, perubahan nilai nominal, manipulasi perangkat/prosedur otorisasi, hingga skema merchant fiktif atau titik kebocoran data (*point of compromise*). Walau detail teknisnya berada di level operasional, intinya adalah bahwa konteks merchant dan proses otorisasi dapat membentuk pola anomali yang menguatkan indikasi *fraud* (Fardiansyah dkk., 2022).

Berdasarkan dimensi di atas, indikator *fraud* pada level data transaksi umumnya direpresentasikan melalui atribut yang menangkap nilai transaksi, waktu transaksi, intensitas/frekuensi transaksi, serta konteks merchant/kanal transaksi karena aspek-aspek tersebut merefleksikan penyimpangan pola dan mekanisme penyalahgunaan yang dijelaskan dalam literatur (Apriwandi dan Herycson, 2022) (Fardiansyah dkk., 2022) (idris, 2020) (Prasetyo, 2021) (Bhatla, 2003).

2.12 Google Colab

Google Colab adalah platform berbasis *cloud* dari *Google* yang memungkinkan pengguna menulis, menjalankan, dan berbagi kode *Python* secara interaktif dalam lingkungan *Jupyter Notebook* tanpa perlu instalasi perangkat lunak khusus. Platform ini menyediakan akses gratis ke sumber daya komputasi seperti GPU dan TPU, yang sangat membantu dalam mempercepat pelatihan model *machine learning* dan analisis data berukuran besar. *Google Colab* mendukung integrasi dengan *Google Drive*, memudahkan penyimpanan dan kolaborasi *real-time* antar pengguna, serta kompatibel dengan berbagai pustaka *Python* populer untuk keperluan *data science* dan AI (Raharjo, Putra, Heristian, dan Napiah, 2025). Dengan kemudahan akses dan kapabilitas komputasi tinggi, *Google Colab* menjadi alat yang sangat berguna bagi peneliti, praktisi, dan mahasiswa dalam mengembangkan dan menguji model-model *machine learning* secara efisien dan kolaboratif.

2.13 Penelitian Terdahulu

Kajian terhadap penelitian terdahulu diperlukan untuk memahami perkembangan metode deteksi penipuan kartu kredit serta mengidentifikasi pendekatan yang paling efektif dalam menangani ketidakseimbangan data. Tabel berikut merangkum studi-studi utama yang relevan, mencakup algoritma yang digunakan, teknik penyeimbangan data, serta performa model yang dilaporkan.

Tabel 2.2. Penelitian Terdahulu

Literatur Referensi	Judul	Model	Performance
(Ding, Liu, Wang, Shi, dan Yu, 2024)	<i>An AutoEncoder Enhanced Light Gradient Boosting Machine Method for Credit Card Fraud Detection</i>	Algoritma: <i>AutoEncoder</i> + <i>LightGBM</i> (AE-LightGBM). Teknik <i>undersampling</i> : <i>Random UnderSampler</i> , <i>NearMiss</i> , <i>Condensed Nearest Neighbour</i> , <i>One-Sided Selection</i> , <i>Edited Nearest Neighbors (ENN)</i> , <i>Instance Hardness Threshold</i> . Teknik <i>oversampling</i> : <i>SMOTE</i> , <i>ADASYN</i> , <i>Borderline-SMOTE</i> .	Menggunakan <i>SMOTE</i> dan <i>AE-LightGBM</i> diperoleh hasil terbaik dengan Recall 94.85%, BCR 97%, AUC 96.83%, dan F1-score 80.27%. Teknik <i>undersampling</i> menghasilkan performa rendah.

Tabel 2.2 Penelitian Terdahulu (Lanjutan)

Literatur Referensi	Judul	Model	Performance
(Siam, Bhowmik, dan Uddin, 2025)	<i>Hybrid Feature Selection Framework for Enhanced Credit Card Fraud Detection Using Machine Learning Models</i>	Algoritma: Random Forest, Extra Trees, XGBoost, Adaboost, CatBoost. Teknik <i>balancing</i> : SMOTE.	Model <i>Extra Trees</i> menghasilkan performa terbaik dengan akurasi 99.97%, F1-score tinggi, serta AUC 98.72% setelah penerapan SMOTE.
(Yadavalli dan Poliseti, 2025)	<i>Optimized Financial Fraud Detection Using SMOTE-Enhanced Ensemble Learning with CatBoost and LightGBM</i>	Algoritma: <i>Logistic Regression</i> , KNN, <i>Decision Tree</i> , Random Forest, XGBoost, CatBoost, LightGBM. Teknik <i>balancing</i> : SMOTE.	SMOTE-enhanced ensemble (CatBoost + LightGBM) mencapai Recall 88%, Precision 68%, dan ROC-AUC 0.9971.
(Salaudeen, Gabi, Muhammad, dan Suru, 2024)	<i>Light Gradient Boosting Machine (LGBM) for Credit Card Fraud in Financial Institution</i>	Algoritma: Logistic Regression, Random Forest, Isolation Forest, MLP, ANN, CNN. Teknik <i>balancing</i> : SMOTE, <i>Random Under-Sampling</i> .	LightGBM menghasilkan performa terbaik dengan Akurasi 96%, Recall 95%, F1-score 96%, dan MCC 93%.
(Gostkowski, Krasnodebski, dan Niedziółka, 2024)	<i>Credit Card Fraud Detection Using Machine Learning Techniques</i>	Algoritma: Decision Tree, ANN, Random Forest. Teknik <i>balancing</i> : Random Undersampling, Random Oversampling, SMOTE.	SMOTE menghasilkan F1-score tertinggi. Random Forest menjadi <i>classifier</i> paling efektif.
(Gupta dkk., 2023)	<i>Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques</i>	Algoritma: KNN, XGBoost, Decision Tree, Random Forest, Logistic Regression. Teknik <i>balancing</i> : <i>Random Oversampling</i> , SMOTE, <i>Undersampling</i> .	XGBoost memberikan performa terbaik dengan <i>Precision</i> dan <i>Accuracy</i> sebesar 0.99 setelah <i>Random Oversampling</i> .

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2.2 Penelitian Terdahulu (Lanjutan)

Literatur Referensi	Judul	Model	Performance
Du, Lv, Wang, dan Guo, (2024)	<i>A Novel Method for Detecting Credit Card Fraud Problems</i>	Algoritma: KNN, AE-XGB, LightGBM, Random Forest. Teknik <i>balancing</i> : SMOTE + CGAN.	AE-XGB-SMOTE-CGAN meningkatkan performa secara signifikan dibanding metode lain.
(Payebi dan Kafhali, 2025)	<i>A Novel Approach Based on XGBoost Classifier and Bayesian Optimization for Credit Card Fraud Detection</i>	Algoritma: XGBoost, LightGBM, AdaBoost, Gradient Boosting, Random Forest, Decision Tree, KNN, Logistic Regression, CatBoost, GaussianNB. Teknik <i>balancing</i> : SMOTE, Random Under-Sampling.	XGBoost dengan <i>Bayesian Optimization</i> dan SMOTE mencapai <i>Accuracy</i> 0.9996, <i>Precision</i> 0.9406, <i>Recall</i> 0.8740, <i>F1-score</i> 0.8740, dan AUC 0.9879.
(F. Zhang, 2023)	<i>Improved Credit Card Fraud Detection Method Based on XGBoost Algorithm</i>	Algoritma: Logistic Regression, Neural Network, XGBoost. Teknik <i>balancing</i> : SMOTE, NearMiss.	XGBoost + SMOTE memberikan performa paling seimbang, sedangkan NearMiss + XGBoost menghasilkan <i>Recall</i> tertinggi.
(Wijaya, Pinaringgi, Zakiyyah, dan Meiliana, 2024)	<i>Comparative Analysis of Machine Learning Algorithms and Data Balancing Techniques for Credit Card Fraud Detection</i>	Algoritma: Logistic Regression, Decision Tree, Random Forest, XGBoost. Teknik <i>balancing</i> : ROS, SMOTE, ADASYN, NCR.	XGBoost + ROS menghasilkan <i>F1-score</i> tertinggi (92.43%), sementara Random Forest + ROS juga menunjukkan performa tinggi.
(Yu, Chang, Hun-Jesus, dan Luo, 2025)	<i>Advanced Supervised Machine Learning Algorithms in Credit Card Fraud Detection</i>	Algoritma: CatBoost, XGBoost, Random Forest. Teknik <i>balancing</i> : SMOTE, ADASYN, Random Undersampling, NearMiss.	CatBoost menunjukkan performa paling seimbang dengan <i>F1-score</i> tertinggi dan ROC terbaik.

Berdasarkan ringkasan pada Tabel 2.2, penelitian terdahulu menunjukkan bahwa deteksi penipuan kartu kredit umumnya lebih efektif ketika menggunakan model berbasis *tree boosting* seperti XGBoost, LightGBM, dan CatBoost, terutama saat dikombinasikan dengan teknik penyeimbangan data pada kondisi *imbalanced*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

class. Mayoritas studi melaporkan bahwa pendekatan *oversampling* seperti SMOTE cenderung memberikan peningkatan yang lebih konsisten pada metrik kelas *fraud*, dibandingkan *undersampling* yang berisiko menghilangkan informasi penting dari kelas mayoritas. Selain itu, beberapa penelitian menguatkan bahwa optimasi model melalui seleksi fitur atau tuning parameter dapat memperbaiki stabilitas dan performa. Temuan-temuan ini memperkuat relevansi penelitian ini untuk membandingkan XGBoost, LightGBM, dan CatBoost dengan integrasi SMOTE, optimasi model, serta analisis interpretabilitas untuk memperoleh kombinasi yang paling efektif dan efisien.



UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

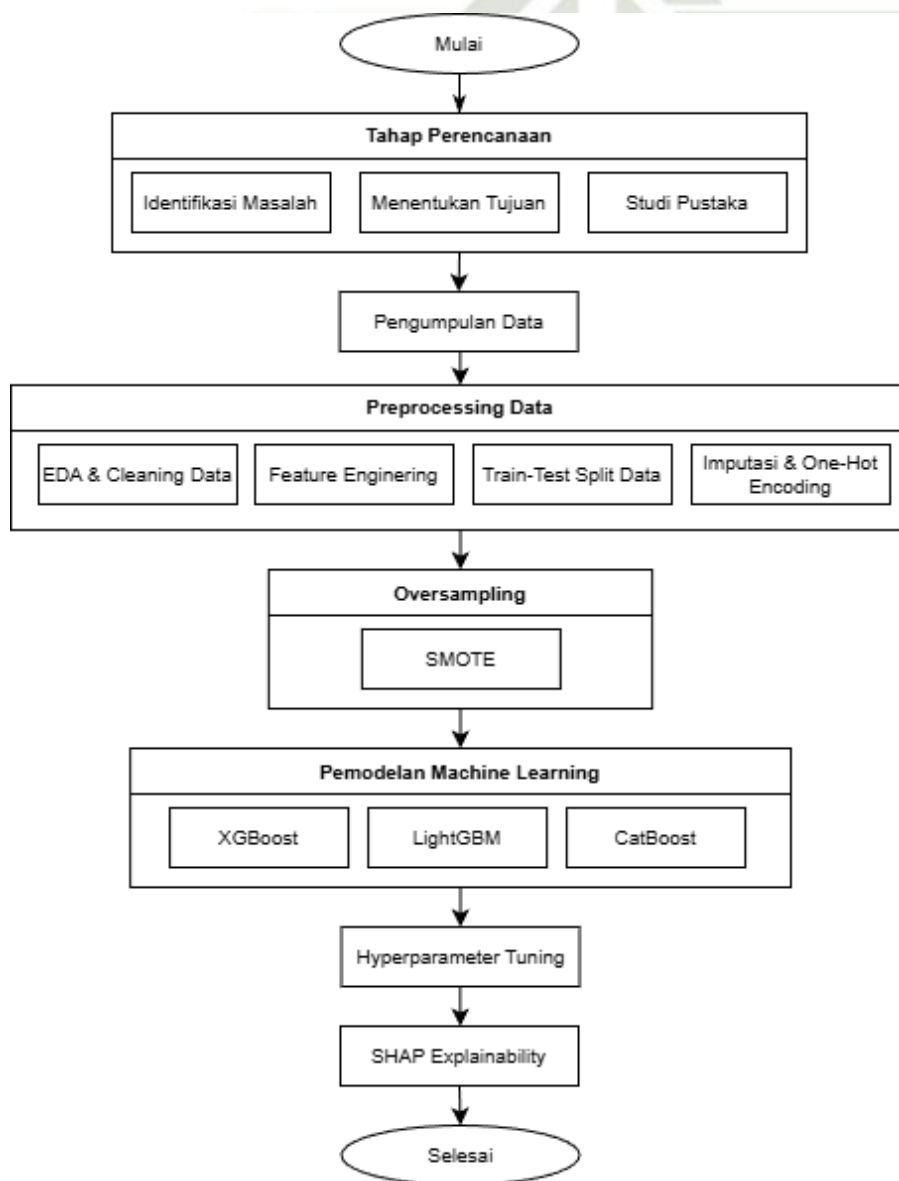
- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak mengutipan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 3 METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

Penelitian ini dilakukan untuk klasifikasi penipuan kartu kredit menggunakan algoritma XGBoost, LightGBM, dan CatBoost agar dapat digunakan sebagai salah satu pendukung keputusan untuk menentukan penipuan terhadap pemakaian kartu kredit. Adapun metodologi penelitian dalam penelitian ini secara garis besar terdiri atas 6 tahapan yang dilakukan. Berikut ini merupakan tahapan penelitian yang dapat dilihat pada Gambar 3.1.



Gambar 3.1. Metodologi Penelitian



3.1.1 Tahap Perencanaan

Tahap perencanaan merupakan tahapan fundamental dan krusial dalam aktivitas penelitian yang bertujuan untuk menjamin bahwa sasaran penelitian dapat ditetapkan secara eksplisit dan tersusun dengan rapi.

3.1.1.1 Identifikasi Masalah

Identifikasi masalah merupakan tahap awal yang dilakukan untuk merumuskan persoalan utama dalam penelitian ini, yaitu klasifikasi penipuan kartu kredit menggunakan dataset transaksi kartu kredit simulasi dari Kaggle yang memuat informasi detail setiap transaksi beserta label '*is_fraud*' (0 = normal, 1 = penipuan). Hasil eksplorasi awal menunjukkan bahwa dataset tersebut memiliki ketidakseimbangan kelas (*imbalanced class*) yang sangat tinggi, di mana proporsi transaksi penipuan hanya kurang dari 1% dari seluruh transaksi sehingga jumlah observasi kelas normal jauh mendominasi. Kondisi ini berpotensi membuat model *machine learning* cenderung bias terhadap kelas mayoritas, menghasilkan nilai akurasi yang tampak tinggi namun gagal mendeteksi sebagian besar transaksi penipuan (nilai *Recall* untuk kelas *fraud* rendah), sehingga penanganan ketidakseimbangan kelas menjadi permasalahan utama yang perlu diatasi dalam penelitian ini.

3.1.1.2 Menentukan Tujuan

Tahap perencanaan merupakan langkah awal yang penting dalam sebuah penelitian untuk memastikan agar tujuan penelitian dapat dirumuskan dengan jelas dan fokus. Penetapan tujuan penelitian yang spesifik dan terukur sangat diperlukan agar penelitian tidak melenceng dari fokus utamanya dan dapat memberikan hasil yang relevan.

3.1.1.3 Studi Literatur

Dilakukan kajian terhadap penelitian terdahulu yang menggunakan algoritma *machine learning*, khususnya XGBoost LightGBM, dan CatBoost, serta pendekatan dalam menangani data tidak seimbang seperti SMOTE, *hyperparameter tuning*, serta analisis interpretabilitas dengan SHAP. Pada proses studi literatur ini dilakukan dengan mengakses beberapa situs web seperti *Science Direct*, IEEE, *Google Scholar* serta sumber-sumber berita aktual lainnya terkait penipuan kartu kredit.

3.1.2 Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data yang bersumber dari dataset publik transaksi kartu kredit simulasi dari Kaggle, yaitu *Credit card Transactions Fraud Detection Dataset* yang dikembangkan oleh Kartik Shenoy. Dataset ini berisi



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

catatan transaksi kartu kredit yang sepenuhnya disintesis dan mencakup transaksi sah maupun penipuan yang dilakukan oleh pemegang kartu terhadap berbagai merchant di wilayah Amerika Serikat bagian barat selama periode kurang lebih satu tahun. Setiap baris data merepresentasikan satu transaksi dengan beragam atribut demografis, geografis, dan perilaku transaksi serta sebuah label biner *is_fraud* yang menunjukkan apakah transaksi tersebut tergolong penipuan atau tidak, sehingga dataset ini dapat dimanfaatkan sebagai sumber data latih dan uji untuk membangun serta mengevaluasi model deteksi penipuan kartu kredit berbasis *machine learning*.

3.1.3 Validasi Data

Validasi pakar dilakukan untuk memastikan bahwa indikator/fitur yang digunakan dalam penelitian deteksi penipuan transaksi kartu kredit relevan secara konseptual dan selaras dengan logika domain, mengingat dataset yang digunakan bersifat publik, anonim, dan *simulated/synthetic*. Proses validasi dilakukan melalui komunikasi tatap muka dan pengisian lembar “Surat Pernyataan Validasi” oleh narasumber ahli, yaitu Lusiawati, S.E., MBA (Dosen/Manajemen Keuangan, UIN Sultan Syarif Kasim Riau), pada 6 Januari 2026 (Lampiran B). Objek validasi difokuskan pada kelayakan indikator/fitur sebagai representasi pola transaksi berisiko (*fraud*) dalam konteks penelitian akademik.

Berdasarkan hasil validasi, narasumber menyatakan bahwa atribut/indikator yang digunakan dalam penelitian ini relevan untuk menggambarkan pola transaksi berisiko. Kelompok fitur yang dinilai relevan meliputi: (1) nilai dan konteks transaksi (misalnya *amt* dan *category*), (2) pola waktu transaksi (misalnya *trans_hour*, *trans_dayofweek*, *day_period*, *is_weekend*), (3) perilaku jangka pendek/indikator *velocity* (misalnya *diff_amt*, *avg_amt_last_7*, *count_last_7*, *category_freq*), (4) lokasi/jarak sebagai sinyal anomali (misalnya *distance_km* dan *state*), serta (5) risiko *merchant* berbasis historis (misalnya *merchant_risk_rate*). Adapun fitur demografis diposisikan sebagai indikator pendukung dan perlu kehati-hatian dalam interpretasi (Lampiran A).

Narasumber juga memberikan catatan bahwa penggunaan fitur *age_group* perlu diperhatikan, khususnya pada kelompok usia ≥ 18 tahun, serta interpretasi terhadap *merchant_risk_rate* perlu disesuaikan karena konteks perilaku dan ekosistem transaksi di Indonesia dapat berbeda dengan karakteristik dataset publik yang digunakan. Dengan demikian, hasil validasi pakar ini memperkuat bahwa pemilihan fitur dalam penelitian telah memiliki dasar konseptual yang memadai, sekaligus menjadi pengingat agar interpretasi hasil dilakukan secara kontekstual sesuai keterbatasan dataset.



3.1.4 Preprocessing Data

Pada tahap *preprocessing data* dilakukan serangkaian langkah untuk menyiapkan dataset sebelum masuk ke proses pemodelan. Mengacu pada rancangan alur penelitian, preprocessing dalam studi ini mencakup empat kegiatan utama, yaitu: (1) *EDA & cleaning data* untuk memahami karakteristik awal dan membersihkan anomali maupun nilai hilang, (2) *Feature engineering* untuk membentuk fitur-fitur turunan yang merepresentasikan pola waktu, demografi, geografis, dan perilaku transaksi, (3) *train-test split* data yang dilakukan secara kronologis menjadi subset latih, validasi, dan uji, serta (4) *imputasi & one-hot encoding* guna menangani nilai hilang dan mengubah variabel kategorikal ke dalam bentuk numerik yang dapat diolah algoritma *machine learning*. Tahap ini bersifat krusial karena menentukan kualitas dan representativitas data yang digunakan sebagai masukan model.

3.1.4.1 Exploratory Data Analysis (EDA)

Tahap *Exploratory Data Analysis* (EDA) dilakukan untuk memperoleh pemahaman awal terhadap struktur data, distribusi variabel, dan potensi pola yang relevan dengan kasus deteksi penipuan. Proses ini mencakup analisis terhadap proporsi kelas pada variabel target *is_fraud*, guna memastikan tingkat ketidakseimbangan data yang akan dihadapi dalam pemodelan.

Selanjutnya, dilakukan analisis distribusi terhadap fitur-fitur temporal seperti jam transaksi (*trans_hour*) dan hari dalam minggu (*trans_dayofweek*) untuk mengidentifikasi kemungkinan adanya pola waktu tertentu yang berasosiasi dengan transaksi *fraud*. Fitur kategorikal seperti *category* juga dianalisis untuk melihat frekuensi serta korelasinya dengan insiden penipuan.

Selain itu, fitur numerik hasil rekayasa seperti *distance_km* dan *age* diamati distribusinya guna mendeteksi *outlier* atau pola khusus yang mungkin relevan dengan perilaku transaksi mencurigakan. Visualisasi dalam bentuk histogram, grafik batang, dan *boxplot* digunakan sebagai alat bantu untuk memperjelas distribusi dan hubungan antarfaktor.

Hasil dari tahap EDA ini tidak hanya digunakan untuk mendukung proses pembersihan dan transformasi data, tetapi juga menjadi dasar dalam perancangan fitur baru dan justifikasi metodologi selanjutnya.

3.1.4.2 Cleaning Data

Pada tahap *cleaning data* dilakukan pembersihan awal terhadap dataset untuk memastikan kelayakan data sebelum *Feature engineering* dan pemodelan. Kegiatan pada tahap ini secara umum mencakup pengecekan kualitas data (seperti keberadaan nilai hilang, duplikasi, dan inkonsistensi dasar) sehingga data yang



3.1.4.3 Feature Engineering

digunakan pada tahap berikutnya berada dalam kondisi lebih bersih dan reliabel.

3.1.4.3 Feature Engineering

Tahap *feature engineering* pada penelitian ini mengadaptasi dan memodifikasi pendekatan yang digunakan dalam penelitian (Alkhozae, 2025), yang menekankan pengayaan fitur temporal, perilaku, dan risiko agregat untuk meningkatkan kemampuan model dalam mendeteksi penipuan transaksi kartu kredit. *Feature engineering* dilakukan untuk mengubah atribut mentah menjadi fitur yang lebih informatif dan relevan, dengan tetap meminimalkan potensi *data leakage*. Secara garis besar, fitur yang dibangun mencakup dimensi temporal, demografis, geografis, perilaku transaksi, dan profil risiko agregat pelanggan maupun merchant.

Dari variabel *trans_date_trans_time* diturunkan fitur berbasis waktu seperti *trans_hour*, *trans_dayofweek*, *is_weekend*, dan *day_period* untuk merepresentasikan pola transaksi berdasarkan jam, hari, dan periode waktu. Tanggal lahir (*dob*) dikonversi menjadi *age* dan *age_group* untuk menangkap perbedaan perilaku antar kelompok umur tanpa menggunakan identitas mentah. Fitur geografis *distance_km* dihitung sebagai jarak geodesik antara pelanggan dan merchant menggunakan koordinat *latitude-longitude*, sehingga merepresentasikan kedekatan lokasi transaksi dengan domisili pelanggan.

Selain itu, dibentuk fitur perilaku historis per kartu dengan mengurutkan transaksi berdasarkan *cc_num* dan *trans_date_trans_time*, kemudian menghitung selisih nilai transaksi dengan transaksi sebelumnya (*diff_amt*) serta statistik rolling tujuh transaksi terakhir (*avg_amt_last_7*, *count_last_7*). Fitur agregat seperti *category_freq* dan *merchant_risk_rate* dihitung dari data pelatihan sebagai frekuensi penggunaan kategori dan rata-rata fraud rate per merchant, lalu diproyeksikan ke data validasi dan uji; merchant baru diisi dengan rata-rata global. Seluruh fitur hasil rekayasa ini, bersama sebagian atribut asli terpilih, digunakan sebagai himpunan fitur akhir dalam proses pemodelan, sedangkan analisis deskriptif dan pengaruhnya terhadap kinerja model dibahas lebih lanjut pada Bab 4.

3.1.4.4 Train-Test Split Data

Setelah dilakukan pengurutan berdasarkan waktu transaksi (*trans_date_trans_time*), dataset dibagi secara kronologis menjadi tiga bagian, yaitu data latih, data validasi, dan data uji dengan proporsi kurang lebih 70%:15%:15%. Data latih digunakan untuk proses pelatihan model, termasuk penerapan SMOTE dan pencarian *hyperparameter*. Data validasi dimanfaatkan untuk memilih kombinasi *hyperparameter* terbaik serta menyetel ambang batas (*threshold*) klasifikasi. Data



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

uji hanya digunakan satu kali pada tahap akhir untuk mengukur kemampuan generalisasi model terhadap transaksi pada periode waktu yang tidak pernah digunakan selama proses pelatihan dan validasi.

3.1.4.5 Imputasi dan One-Hot Encoding

Sebelum pemodelan, dilakukan imputasi untuk menangani nilai hilang dan transformasi variabel kategorikal. Nilai hilang pada fitur numerik diisi menggunakan nilai median yang dihitung dari data latih, sedangkan nilai hilang pada fitur kategorikal diisi menggunakan modus dari data latih. Statistik imputasi ini kemudian diterapkan secara konsisten pada data validasi dan data uji. Selanjutnya, fitur kategorikal yang digunakan dalam pemodelan diubah menjadi representasi numerik menggunakan teknik one-hot encoding dengan skema yang sama untuk data latih, validasi, dan uji, sehingga struktur fitur yang terbentuk tetap konsisten di seluruh himpunan data.

3.1.5 Oversampling Data

Metode *oversampling* yang disebut *Synthetic Minority Over-sampling Technique* (SMOTE) digunakan untuk mengatasi ketidakseimbangan data. Dengan memanfaatkan tetangga terdekatnya berdasarkan jarak *Euclidean* untuk membuat data sintetis dari kelas minoritas. SMOTE meningkatkan jumlah sampel karena data sintetis dibuat menggunakan fitur yang sudah ada sebelumnya, sehingga terlihat mirip dengan data asli (R. Chen dkk., 2023).

Untuk mengatasi masalah ketidakseimbangan data, teknik SMOTE akan diterapkan untuk menyeimbangkan kelas data. Dengan memanfaatkan tetangga terdekatnya berdasarkan jarak *Euclidean* untuk membuat data sintetis dari kelas minoritas. SMOTE meningkatkan jumlah sampel karena data sintetis dibuat menggunakan fitur yang sudah ada sebelumnya, sehingga terlihat mirip dengan data asli.

3.1.6 Pemodelan Machine Learning

Pada tahap pemodelan *machine learning*, data yang telah diproses melalui tahapan pembersihan dan penyeimbangan selanjutnya digunakan untuk membangun model prediksi penipuan kartu kredit. Algoritma yang digunakan dalam penelitian ini Adalah XGBoost, LightGBM, dan CatBoost, ketiganya merupakan metode *boosting* yang mampu menangani data kompleks serta pola *non-linier* pada transaksi kartu kredit.

Pada tahap ini, masing-masing algoritma dilatih menggunakan data latih yang telah dipersiapkan pada tahap sebelumnya, hingga memperoleh pola yang optimal dalam membedakan transaksi penipuan dan bukan penipuan. Selama pelati-

han model, dilakukan pemantauan metrik performa seperti *Accuracy*, *Precision*, *Recall*, *F1-score*, dan AUC-ROC untuk memastikan bahwa model dapat mengidentifikasi transaksi penipuan secara efektif.

Tahap pemodelan ini sangat krusial, karena hasil yang diperoleh dari model machine learning akan digunakan sebagai dasar dalam evaluasi dan analisis efektivitas deteksi penipuan kartu kredit pada penelitian.

3.1.7 Hyperparameter Tuning

Hyperparameter tuning, yaitu usaha untuk menemukan kombinasi terbaik dari hyperparameter guna meningkatkan performa model. Tujuan dari proses ini adalah untuk memaksimalkan akurasi, kemampuan generalisasi, atau efektivitas model dengan mengatur parameter-parameter yang bisa dikonfigurasi. *Grid Search Cross-Validation* (GridSearchCV) merupakan salah satu metode yang digunakan untuk menentukan kombinasi parameter terbaik dalam suatu model. Metode ini bekerja dengan mengevaluasi setiap kemungkinan kombinasi parameter secara sistematis, lalu melakukan validasi terhadap masing-masing kombinasi tersebut.

Untuk memperoleh konfigurasi model yang optimal, dilakukan *hyperparameter tuning* menggunakan GridSearchCV dengan skema *TimeSeriesSplit*, sehingga validasi dilakukan secara berurutan mengikuti dimensi waktu pada data transaksi. GridSearchCV dijalankan pada *pipeline* yang berisi SMOTE dan masing-masing algoritma (XGBoost, LightGBM, dan CatBoost), dengan himpunan *hyperparameter* seperti pada Tabel 3.1

Selanjutnya, dilakukan penyetelan *threshold* (*decision threshold tuning*) menggunakan data validasi. Untuk tiap model terbaik, nilai probabilitas keluaran (*predict_proba*) pada kelas *fraud* dihitung, kemudian beberapa nilai *threshold* diuji (misalnya pada rentang 0,1–0,9) untuk mencari titik potong yang memberikan *F1-score* terbaik pada data validasi. *Threshold* optimal ini kemudian disimpan dan digunakan pada tahap evaluasi akhir di data uji, sehingga prediksi kelas tidak hanya bergantung pada *threshold default* 0,5, tetapi telah disesuaikan dengan karakteristik ketidakseimbangan data.

Tabel 3.1. Nilai Terbaik untuk *Hyperparameter*

Algoritma	Hyperparameter	Kegunaan Hyperparameter
XGBoost	<i>n_estimators</i>	Banyaknya tree yang digunakan untuk proses klasifikasi
	<i>max_depth</i>	Kedalaman maksimum tree yang dapat dibentuk
	<i>reg_alpha</i>	Regularisasi L1 pada bobot leaf; mengurangi kompleksitas efektif model dan membantu menekan overfitting

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 3.1 Nilai Terbaik untuk *Hyperparameter* (Lanjutan)

Algoritma	Hyperparameter	Kegunaan Hyperparameter
	<i>eta (learning_rate)</i>	Membantu mempersingkat langkah dalam pembaruan model
	<i>subsample</i>	Rasio instance dari data latih yang digunakan pada setiap pohon
	<i>colsample_bytree</i>	Persentase fitur yang diambil per pohon; meningkatkan diversifikasi pohon dan kemampuan generalisasi
LightGBM	<i>max_depth</i>	Membatasi kedalaman maksimum pohon untuk mengontrol kompleksitas model
	<i>num_leaves</i>	Menentukan jumlah maksimum leaf per pohon; nilai terlalu besar membuat model sangat fleksibel dan berisiko <i>overfitting</i>
	<i>min_child_samples</i>	Jumlah minimum sampel pada satu leaf; nilai lebih besar menghasilkan keputusan yang lebih stabil
	<i>learning_rate</i>	Mengatur kontribusi setiap pohon baru terhadap model akhir
	<i>n_estimators</i>	Jumlah iterasi <i>boosting</i> ; memengaruhi performa dan waktu komputasi
	<i>reg_alpha</i>	Regularisasi L1 untuk mendorong bobot <i>leaf</i> menjadi lebih <i>sparse</i> dan mengendalikan <i>overfitting</i>
	<i>subsample</i>	Proporsi data latih yang digunakan untuk setiap pohon guna meningkatkan generalisasi
	<i>colsample_bytree</i>	Proporsi fitur yang digunakan per pohon untuk mengurangi korelasi antar pohon
CatBoost	<i>depth</i>	Mengatur kedalaman maksimum pohon; nilai lebih besar meningkatkan kompleksitas model
	<i>learning_rate</i>	Mengontrol besar langkah pembaruan pada setiap iterasi <i>boosting</i>
	<i>iterations</i>	Menentukan jumlah pohon <i>boosting</i> ; semakin banyak iterasi meningkatkan kapasitas model
	<i>l2_leaf_reg</i>	Regularisasi L2 pada nilai <i>leaf</i> untuk menghaluskan prediksi dan menekan <i>overfitting</i>
	<i>bagging_temperature</i>	Mengatur tingkat keacakan pemilihan sampel; nilai lebih tinggi meningkatkan variasi data

3.1.8 Evaluasi Model *Machine Learning*

Kinerja setiap model dievaluasi secara komprehensif menggunakan data uji untuk menilai kemampuan generalisasinya pada data baru. Titik awal evaluasi adalah confusion matrix, yang memetakan hasil prediksi ke dalam empat kategori: *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN).

Dari matriks ini, dihitung serangkaian metrik performa. Mengingat sifat data yang tidak seimbang, metrik seperti *Accuracy*, *Precision*, *Recall*, dan F1-score menjadi fokus utama untuk mengukur efektivitas model dalam mengidentifikasi kelas minoritas (*fraud*). Selain itu, metrik AUC-ROC digunakan untuk menilai kemampuan model secara keseluruhan dalam membedakan antara transaksi normal dan *fraud*. Hasil perbandingan metrik-metrik inilah yang menjadi dasar untuk menentukan model paling optimal.

3.1.8.1 Analisis Interpretabilitas dengan SHAP

Sebagai bagian dari pendekatan *Explainable Artificial Intelligence* (XAI), penelitian ini menggunakan *SHapley Additive exPlanations* (SHAP) untuk menganalisis interpretabilitas model dengan kinerja terbaik pada tingkat global. SHAP digunakan untuk menjelaskan seberapa besar kontribusi relatif masing-masing fitur terhadap output model secara keseluruhan, sehingga tidak hanya diketahui model mana yang paling baik, tetapi juga fitur apa yang paling berperan dalam membedakan transaksi *fraud* dan *non-fraud*.

Implementasi SHAP dilakukan dengan memanfaatkan *TreeExplainer* pada model terbaik, kemudian menghitung nilai SHAP untuk sekelompok sampel data (misalnya himpunan validasi atau uji yang telah melalui tahapan *preprocessing* dan *feature engineering*). Nilai SHAP yang dihasilkan kemudian divisualisasikan dalam bentuk *summary plot* (baik *bar plot* maupun *beeswarm plot*) untuk memperoleh gambaran global mengenai urutan kepentingan fitur serta arah pengaruhnya terhadap peningkatan atau penurunan risiko *fraud*. Selain itu, digunakan pula *dependence plot* pada beberapa fitur kunci hasil *feature engineering*, seperti *merchant_risk_rate*, *category_freq*, *diff_amt*, dan *distance_km*, guna melihat hubungan antara nilai fitur dengan besarnya kontribusi SHAP pada berbagai sampel.

Analisis SHAP ini tidak mengubah struktur maupun parameter model, melainkan berfungsi sebagai lapisan penjelas untuk menginterpretasikan perilaku model yang telah terlatih. Informasi mengenai fitur-fitur yang paling berpengaruh dan pola kontribusinya digunakan sebagai dasar untuk mengidentifikasi karakteristik umum transaksi yang berisiko penipuan, yang selanjutnya dibahas lebih rinci pada Bab 4.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan rangkaian keseluruhan tahapan penelitian yang meliputi eksplorasi data, prapemrosesan, pemodelan, evaluasi kinerja, dan analisis interpretabilitas, diperoleh beberapa kesimpulan utama sebagai berikut:

1. Pada skenario *baseline* tanpa SMOTE dan tanpa *hyperparameter tuning*, XGBoost, LightGBM, dan CatBoost sudah menghasilkan *Accuracy* dan AUC-ROC yang tinggi, tetapi kinerja pada kelas *fraud* masih bermasalah karena ketidakseimbangan kelas yang ekstrem. Hal ini tercermin dari *Precision* kelas *fraud* yang relatif rendah. Setelah diterapkan SMOTE dengan rasio 10% pada data latih dan diikuti GridSearchCV serta penyesuaian *decision threshold*, model terbaik pada data uji mencapai *F1-score* sekitar 0,90, AUC-ROC di atas 0,996, dan *Accuracy* 0,9993–0,9994. Dalam konfigurasi akhir tersebut, CatBoost memperoleh *F1-score* dan *Recall* tertinggi, sedangkan XGBoost memberikan *Precision* tertinggi dengan kinerja yang tetap kompetitif, sehingga kombinasi SMOTE dan *hyperparameter tuning* terbukti meningkatkan kualitas deteksi *fraud*.
2. Analisis *Shapley Additive Explanations* (SHAP) menunjukkan bahwa ketiga model mengandalkan kombinasi fitur temporal, perilaku, dan risiko historis dalam memprediksi *fraud*. Fitur *amt* dan *day_period_night* secara konsisten menjadi kontributor utama, disusul fitur-fitur seperti *merchant_risk_rate*, *category_freq*, dan *avg_amt_last_7*. SHAP memungkinkan pemetaan yang jelas mengenai fitur yang mendorong peningkatan ataupun penurunan probabilitas *fraud*, sehingga interpretabilitas dan transparansi model tercapai.
3. Untuk deteksi penipuan kartu kredit pada data yang sangat tidak seimbang, penelitian ini merekomendasikan *pipeline* berbasis pembagian data temporal, *data cleaning* dan *feature engineering* terarah, imputasi berbasis statistik data latih, *one-hot encoding* untuk fitur kategorikal, SMOTE dengan rasio 10% pada data latih, serta pelatihan model boosting yang dioptimalkan menggunakan GridSearchCV dan penyesuaian *decision threshold*. Dalam kerangka ini, CatBoost + SMOTE 10% + *tuning* sesuai untuk skenario yang memprioritaskan *F1-score* dan *Recall fraud*, sedangkan XGBoost + SMOTE 10% + *tuning* lebih tepat untuk kasus yang menekankan *Precision* tinggi dan pengendalian *false positives*, dengan LightGBM sebagai alternatif kompromi yang lebih ringan secara komputasi.



5.2 Saran

Dalam penelitian selanjutnya, disarankan untuk memperluas eksplorasi teknik penanganan ketidakseimbangan kelas dan optimasi model, misalnya dengan membandingkan variasi metode oversampling lain (*ADASYN*, *Borderline-SMOTE*, atau kombinasi *SMOTE* dengan *under-sampling*) serta pendekatan tuning yang lebih maju seperti *RandomizedSearchCV* atau optimasi Bayesian. Selain itu, model dan temuan empiris pada penelitian ini dapat dijadikan landasan awal untuk pengembangan prototipe sistem deteksi penipuan kartu kredit berbasis *machine learning* yang terintegrasi dengan lingkungan operasional, misalnya dalam bentuk *dashboard* pemantauan transaksi secara near real-time atau modul pendukung keputusan bagi analisis *fraud*. Dengan demikian, kontribusi penelitian tidak hanya berhenti pada level evaluasi model, tetapi juga dapat diarahkan menuju implementasi sistem yang aplikatif di dunia nyata.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



DAFTAR PUSTAKA

- A., S., dan R., S. (2023, 6). A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, 7, 100230. doi: 10.1016/j.dajour.2023.100230
- Adepoju, O., dan Wosowei, J. (2019). lawte, s., & jaiman, h.(2019). comparative evaluation of credit card fraud detection using machine learning techniques. Dalam *2019 global conference for advancement in technology (gcatt)*. doi (Vol. 10).
- Alriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., ... Eshun, J. (2023, 3). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal*, 6, 100163. doi: 10.1016/j.dajour.2023.100163
- Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., ... Herrera, F. (2023, 11). Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion*, 99, 101805. doi: 10.1016/j.inffus.2023.101805
- Alkhozae, M. (2025, 12). An explainable credit card fraud detection model using machine learning and deep learning approaches. *Journal of Applied Data Sciences*, 6, 2838-2857. doi: 10.47738/jads.v6i4.962
- Andrade-Arenas, L., dan Yactayo-Arias, C. (2025, 5). Comparative analysis of machine learning models for credit card fraud detection using smote for class imbalance. *International Journal of Safety and Security Engineering*, 15, 893-901. doi: 10.18280/ijssse.150504
- Apriwandi, A., dan Herycson, H. (2022, 9). Cyber crime dan fraud kartu kredit dan kartu debit: Perspektif akuntansi. *JUEB : Jurnal Ekonomi dan Bisnis*, 1, 111-124. doi: 10.57218/jueb.v1i3.277
- Babu, E., Maliakal, J. J., V, N. T., dan Babu, D. (2024, 12). Performance comparison of xgboost and catboost algorithm in credit card fraud detection with streamlit-based web application. Dalam *2024 international conference on advancement in renewable energy and intelligent systems (areis)* (hal. 1-6). IEEE. doi: 10.1109/AREIS62559.2024.10893617
- Bhatla, T. P. (2003). *Understanding credit card frauds*. TATA Consultancy Services.
- Bifarir, O. O. (2023, 5). Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification. *PLOS ONE*, 18, e0284315. doi: 10.1371/journal.pone.0284315



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Chakraborty, C., Bhattacharya, M., Pal, S., dan Lee, S.-S. (2024). From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare. *Current Research in Biotechnology*, 7, 100164. doi: 10.1016/j.crbiot.2023.100164
- Chang, V., Ali, B., Golightly, L., Ganatra, M. A., dan Mohamed, M. (2024, 8). Investigating credit card payment fraud with detection methods using advanced machine learning. *Information*, 15, 478. doi: 10.3390/info15080478
- Chen, R., Zhang, S., Li, J., Guo, D., Zhang, W., Wang, X., ... Wang, X. (2023, 3). A study on predicting the length of hospital stay for chinese patients with ischemic stroke based on the xgboost algorithm. *BMC Medical Informatics and Decision Making*, 23, 49. doi: 10.1186/s12911-023-02140-4
- Chen, Y., dan Han, X. (2021, 1). Catboost for fraud detection in financial transactions. Dalam *2021 ieee international conference on consumer electronics and computer engineering (iccece)* (hal. 176-179). IEEE. doi: 10.1109/ICCECE51280.2021.9342475
- Consumer sentinel network data book 2021*. (2022, 2). Retrieved from <https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2021>
- Detthamrong, U., Chansanam, W., Boongoen, T., dan Iam-On, N. (2024, 9). Enhancing fraud detection in banking using advanced machine learning techniques. *International Journal of Economics and Financial Issues*, 14, 177-184. doi: 10.32479/ijefi.16613
- Dev, V. A., dan Eden, M. R. (2019). Gradient boosted decision trees for lithology classification. Dalam (hal. 113-118). doi: 10.1016/B978-0-12-818597-1.50019-9
- Ding, L., Liu, L., Wang, Y., Shi, P., dan Yu, J. (2024, 10). An autoencoder enhanced light gradient boosting machine method for credit card fraud detection. *PeerJ Computer Science*, 10, e2323. doi: 10.7717/peerj-cs.2323
- Du, H. C., Lv, L., Wang, H., dan Guo, A. (2024, 3). A novel method for detecting credit card fraud problems. *PLoS ONE*, 19. Retrieved from <https://doi.org/10.1371/journal.pone.0294537.t004> doi: 10.1371/journal.pone.0294537
- Fadzail, N. F., Zali, S. M., Mid, E. C., dan Bakar, N. N. A. (2025, 12). Hyperparameter optimization using automated machine learning based artificial neural network for fault detection in wind turbine generator. *Next Research*, 2, 100785. doi: 10.1016/j.nexres.2025.100785
- Faraji, Z. (2022, 2). A review of machine learning applications for credit card fraud

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

detection with a case study. *SEISENSE Journal of Management*, 5, 49-59. doi: 10.33215/sjom.v5i1.770

Fardiansyah, H., Satory, A., dan Dewi, R. (2022). *Cyber crime paling populer pada era digital* (R. R. Rerung, Ed.). Media Sains Indonesia.

Fraud detection in indonesia's financial sector. (2025, 9).

GeeksforGeeks. (2025, 7). *Catboost in machine learning*. Retrieved from <https://www.geeksforgeeks.org/machine-learning/catboost-ml/>

Gostkowski, M., Krasnodebski, A., dan Niedziółka, A. (2024). Credit card fraud detection using machine learning techniques. *European Research Studies Journal*, 27(2), 571–585.

Guo, P., Yang, Y., Guo, W., dan Shen, Y. (2024, 7). A fair contribution measurement method for federated learning. *Sensors*, 24, 4967. doi: 10.3390/s24154967

Gupta, P., Varshney, A., Khan, M. R., Ahmed, R., Shuaib, M., dan Alam, S. (2023). Unbalanced credit card fraud detection data: A machine learning-oriented comparative study of balancing techniques. *Procedia Computer Science*, 218, 2575-2584. doi: 10.1016/j.procs.2023.01.231

Gyamfi, N. K., dan Abdulai, J.-D. (2020, 11). Bank fraud detection using support vector machine. Dalam *2018 ieee 9th annual information technology, electronics and mobile communication conference (iemcon)* (hal. 37-41). IEEE. doi: 10.1109/IEMCON.2018.8614994

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

Hasnaningtyas, N., dan Dewayanto, T. (2023, 11). Financial fraud detection and machine learning algorithm (unsupervised learning): Systematic literature review. *Jurnal Riset Akuntansi Dan Bisnis Airlangga*, 8, 1521-1542. doi: 10.20473/jraba.v8i2.49927

Idris, H. D. P. (2020, 12). Kebijakan dalam penanggulangan penyalahgunaan kartu kredit. *LEX PRIVATUM*.

Irawan, H. (2025, 9). Deteksi fraud kartu kredit dengan logistic regression, random forest, dan gradient boosting. *Jurnal Informatika Terpadu*, 11, 92-97. doi: 10.54914/jit.v11i2.1826

Karo, I. M. K. (2020). *Implementasi metode xgboost dan feature importance untuk klasifikasi pada kebakaran hutan dan lahan* (Vol. 1; Tech. Rep.).

Liang, Y., Li, S., Yan, C., Li, M., dan Jiang, C. (2021, 1). Explaining the black-box model: A survey of local interpretation methods for deep neural networks. *Neurocomputing*, 419, 168-182. doi: 10.1016/j.neucom.2020.08.011



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Liu, L., Li, Y., Gruyer, D., dan Tu, M. (2025, 9). Non-linear relationship between built environment and active travel: a hybrid model considering spatial heterogeneity. *International Journal of Transportation Science and Technology*, 19, 175-192. doi: 10.1016/j.ijtst.2024.08.008
- Lundberg, S. M., dan Lee, S.-I. (2020). A unified approach to interpreting model predictions. Dalam I. Guyon dkk. (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
- M, P., K., P. R., R., N., AboRas, K. M., dan Youssef, A. (2024, 6). Application of an ensemble catboost model over complex dataset for vehicle classification. *PLOS ONE*, 19, e0304619. doi: 10.1371/journal.pone.0304619
- Majumder, T. (2024, 12). Financial fraud detection for credit card using xgboost & smote. *Nanotechnology Perceptions*, 32-50. doi: 10.62441/nano-ntp.vi.3425
- Mathivanan, N. M. N., Md.Ghani, N. A., dan Janor, R. M. (2019, 11). A comparative study on dimensionality reduction between principal component analysis and k-means clustering. *Indonesian Journal of Electrical Engineering and Computer Science*, 16, 752. doi: 10.11591/ijeecs.v16.i2.pp752-758
- Matsuo, Y., LeCun, Y., Sahani, M., Doina, P., David, S., Sakai, M., ... Matsumoto, J. (2020). Machine learning: A review of learning types. *Neural Networks*, 7(1), 267–275. Retrieved from <https://doi.org/10.20944/preprints202007.0230.v1> doi: 10.20944/preprints202007.0230.v1
- Molnar, C. (2020). *Interpretable machine learning ; a guide for making black box models explainable*. Leanpub.
- Noh, J., dan Kang, S.-S. (2025, 8). Evaluation of prediction models for slope stability on road cut slopes using xgboost, lightgbm, and catboost. *KSCE Journal of Civil Engineering*, 100379. doi: 10.1016/j.kscej.2025.100379
- Noviandy, T. R., Idroes, G. M., Hardi, I., Afjal, M., dan Ray, S. (2024, 5). A model-agnostic interpretability approach to predicting customer churn in the telecommunications industry. *Infolitika Journal of Data Science*, 2, 34-44. doi: 10.60084/ijds.v2i1.199
- Ogunsanya, M., Isichei, J., dan Desai, S. (2023, 8). Grid search hyperparameter tuning in additive manufacturing processes. *Manufacturing Letters*, 35, 1031-1042. doi: 10.1016/j.mfglet.2023.08.056
- Ok receives 153,000 financial scam reports. (2025, 6). *TEMPO*. Retrieved from <https://en.tempo.co/read/2021366/ojk-receives-153000>



Hak Cipta Dilindungi Undang-Undang

Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

-financial-scam-reports?

- Okagbue, E. F., Ezeachikulo, U. P., Akintunde, T. Y., Tsakuwa, M. B., Ilokanulo, S. N., Obiasoanya, K. M., ... Ouattara, C. A. T. (2023). A comprehensive overview of artificial intelligence and machine learning in education pedagogy: 21 years (2000–2021) of research indexed in the scopus database. *Social Sciences & Humanities Open*, 8, 100655. doi: 10.1016/j.ssaho.2023.100655
- Oluwatoyin, J. A., dan Akinola, S. (2024, 4). Real time credit card fraud detection and reporting system using machine learning. *European Journal of Computer Science and Information Technology*, 12, 36-56. doi: 10.37745/ejcsit.2013/vol12n43656
- Oztornaci, B., Ata, B., dan Kartal, S. (2024, 6). Analysing household food consumption in turkey using machine learning techniques. *Agris on-line Papers in Economics and Informatics*, 16, 97-105. doi: 10.7160/aol.2024.160207
- Ponzone, I., Prosper, J. A. P., dan Campillo, N. E. (2023, 11). Explainable artificial intelligence: A taxonomy and guidelines for its application to drug discovery. *WIREs Computational Molecular Science*, 13. doi: 10.1002/wcms.1681
- Pozzolo, A. D., Boracchi, G., Caelen, O., Alippi, C., dan Bontempi, G. (2018, 8). Credit card fraud detection: A realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks and Learning Systems*, 29, 3784-3797. doi: 10.1109/TNNLS.2017.2736643
- Prasetyo, S. N. (2021). *Rumusan pengaturan credit card fraud dalam hukum pidana indonesia ditinjau dari asas legalitas* (Vol. 24; Tech. Rep.). Retrieved from <http://tekno>.
- Raghavan, P., dan Gayar, N. E. (2019, 12). Fraud detection using machine learning and deep learning. Dalam *2019 international conference on computational intelligence and knowledge economy (iccike)* (hal. 334-339). IEEE. doi: 10.1109/ICCIKE47802.2019.9004231
- Raharjo, M., Putra, J. L., Heristian, S., dan Napiyah, M. (2025, 5). Sentiment analysis of video editing applications using support vector machine on google colab. *JIKA (Jurnal Informatika)*, 9, 192. doi: 10.31000/jika.v9i2.13699
- Rajaan, M. A. K., Sakib, S., Fahad, N. M., Mamun, A. A., Rahman, M. A., Shatabda, S., dan Mukta, M. S. H. (2024, 6). A systematic review of hyperparameter optimization techniques in convolutional neural networks. *Decision Analytics Journal*, 11, 100470. doi: 10.1016/j.dajour.2024.100470
- RAMADHAN, M. M., SITANGGANG, I. S., NASUTION, F. R., dan GHIFARI, A. (2021, 10). Parameter tuning in random forest based on grid search method

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

for gender classification based on voice frequency. *DEStech Transactions on Computer Science and Engineering*. doi: 10.12783/dtcse/cece2017/14611

Ramírez-Sanz, J. M., Maestro-Prieto, J.-A., Álar Arnaiz-González, dan Bustillo, A. (2023, 12). Semi-supervised learning for industrial fault detection and diagnosis: A systemic review. *ISA Transactions*, 143, 255-270. doi: 10.1016/j.isatra.2023.09.027

Roshan, K., dan Zafar, A. (2021, 9). Utilizing xai technique to improve autoencoder based model for computer network anomaly detection with shapley additive explanation(shap). *International journal of Computer Networks & Communications*, 13, 109-128. doi: 10.5121/ijenc.2021.13607

Saeed, W., dan Omlin, C. (2023, 3). Explainable ai (xai): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, 263, 110273. doi: 10.1016/j.knosys.2023.110273

Salaudeen, L. G., Gabi, D., Muhammad, G., dan Suru, H. U. (2024). Light gradient boosting machine (lgbm) for credit card fraud in financial institution. *Direct Research Journal of Engineering and Information Technology*, 12, 19-34. Retrieved from <https://doi.org/10.26765/DRJEIT17933661> doi: 10.26765/DRJEIT17933661

Sankar, S., Potti, A., Chandrika, G. N., dan Ramasubbareddy, S. (2022, 2). Thyroid disease prediction using xgboost algorithms. *Journal of Mobile Multimedia*. doi: 10.13052/jmm1550-4646.18322

Scarano, A., Sadeghi, M., Mauriello, F., Riccardi, M. R., Aghabayk, K., dan Montella, A. (2025, 7). Cyclist crash severity modeling: A hybrid approach of xgboost-shap and random parameters logit with heterogeneity in means and variances. *Journal of Safety Research*, 93, 373-398. doi: 10.1016/j.jsr.2025.04.003

Shenoy, K. (2020). *Credit card transactions fraud detection dataset*. Kaggle. Retrieved from <https://www.kaggle.com/datasets/kartik2112/fraud-detection>

Shi, S., Luo, W., dan Pau, G. (2025, 6). An attention-based balanced variational autoencoder method for credit card fraud detection. *Applied Soft Computing*, 177, 113190. doi: 10.1016/j.asoc.2025.113190

Shobayo, O., Zachariah, O., Odusami, M. O., dan Ogunleye, B. (2023, 8). Prediction of stroke disease with demographic and behavioural data using random forest algorithm. *Analytics*, 2, 604-617. doi: 10.3390/analytics2030034

Siam, A. M., Bhowmik, P., dan Uddin, M. P. (2025, 7). Hybrid feature selection framework for enhanced credit card fraud detection using machine learning



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- models. *PLOS One*, 20, e0326975. doi: 10.1371/journal.pone.0326975
- Sun, M., Kraus, T., Eilts, J., dan Garus, C. (2025, 12). Machine learning for predicting sense of place among hybrid workers: A shap-based analysis of key factors. *Wellbeing, Space and Society*, 9, 100328. doi: 10.1016/j.wss.2025.100328
- Survei fico. (2025, 3). Retrieved from <https://investors.fico.com/news-releases/news-release-details/fico-survey-1-4-indonesian-consumers-report-losing-money-scams/>
- Taha, A. A., dan Malebary, S. J. (2020). An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. *IEEE Access*, 8, 25579-25587. doi: 10.1109/ACCESS.2020.2971354
- Tayebi, M., dan Kafhali, S. E. (2025, 12). A novel approach based on xgboost classifier and bayesian optimization for credit card fraud detection. *Cyber Security and Applications*, 3, 100093. doi: 10.1016/j.csa.2025.100093
- Ukwaththa, J., Herath, S., dan Meddage, D. (2024, 12). A review of machine learning (ml) and explainable artificial intelligence (xai) methods in additive manufacturing (3d printing). *Materials Today Communications*, 41, 110294. doi: 10.1016/j.mtcomm.2024.110294
- Valencia-Arias, A., Coronado, M. H. V., Medina, M. L. C., Garcia, M. J. M., Arias, A. R. R., dan Tirado, J. I. S. (2025, 12). Research trends in the application of machine learning in sustainability practices based on a bibliometric analysis. *Sustainable Futures*, 10, 100987. doi: 10.1016/j.sftr.2025.100987
- Verwaeren, J., der Weeën, P. V., dan Baets, B. D. (2022, 6). A search grid for parameter optimization as a byproduct of model sensitivity analysis. *Applied Mathematics and Computation*, 261, 8-27. doi: 10.1016/j.amc.2015.03.064
- Vincent, F.-L., Peter, H., Riashat, I., Marc, G. B., dan Joelle, P. (2018, 12). An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning*, 11, 219-354. doi: 10.1561/22000000071
- Wijaya, M. G., Pinaringgi, M. F., Zakiyyah, A. Y., dan Meiliana. (2024). Comparative analysis of machine learning algorithms and data balancing techniques for credit card fraud detection. *Procedia Computer Science*, 245, 677-688. doi: 10.1016/j.procs.2024.10.294
- Xiao, Y., Tan, L., dan Liu, J. (2025, 3). Application of machine learning model in fraud identification: A comparative study of catboost, xgboost and lightgbm. doi: 10.20944/preprints202503.1199.v1
- Yadavalli, R., dan Polisetti, R. (2025, 3). Optimized financial fraud detection using smote-enhanced ensemble learning with catboost and lightgbm. Dalam 2025



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

international conference on visual analytics and data visualization (icvadv) (hal. 1554-1562). IEEE. doi: 10.1109/ICVADV63329.2025.10961683

Ya, S., Chang, V., Hunh, G., Jesus, V., dan Luo, J. (2025). Advanced supervised machine learning algorithms in credit card fraud detection. Dalam *Proceedings of the 7th international conference on finance, economics, management and it business* (hal. 126-138). SCITEPRESS - Science and Technology Publications. doi: 10.5220/0013485400003956

Zandieh, M., Kazemi, A., dan Ahmadi, M. (2021, 10). A comprehensive insight into the application of machine learning approaches in predicting the separation efficiency of hydrocyclones. *Desalination and Water Treatment*, 236, 123-143. doi: 10.5004/dwt.2021.27691

Zhang, F. (2023, 3). Improved credit card fraud detection method based on xgboost algorithm. *BCP Business & Management*, 38, 2888-2895. doi: 10.54691/bcpbm.v38i.4206

Zhang, Z. (2016, 6). Introduction to machine learning: k-nearest neighbors. *Annals of Translational Medicine*, 4, 218-218. doi: 10.21037/atm.2016.03.37



UIN SUSKA RIAU



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN A

SURAT PERNYATAAN VALIDASI PAKAR

SURAT PERNYATAAN VALIDASI

Saya yang bertanda tangan di bawah ini:

Nama : LUSIAWATI, SE. MBA
 Jabatan/Bidang : DOSEN / MANAJEMEN KEUANGAN
 Instansi : FAKULTAS UIN SUSKA RIAU

Dengan ini menyatakan bahwa saya telah memberikan validasi pakar terhadap penelitian Tugas Akhir yang disusun oleh:

Nama Mahasiswa : Amelia Rahmadani
 Program Studi : Sistem Informasi
 Fakultas/Universitas : Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau
 Judul Tugas Akhir : Deteksi Penipuan Transaksi Kartu Kredit Menggunakan Algoritma Boosting (XGBoost, LightGBM, dan CatBoost) dengan SMOTE, Hyperparameter Tuning, Threshold Tuning, serta Interpretabilitas SHAP

Objek validasi: Kelayakan indikator/fitur yang digunakan dalam penelitian deteksi fraud berbasis machine learning.

Catatan: Dataset bersifat publik, anonim, dan simulated/synthetic, validasi ini tidak menggunakan data internal lembaga tertentu.

A. Pernyataan Validasi (centang)

Berdasarkan penjelasan dan lampiran yang diberikan, saya menyatakan bahwa:

1. Atribut/indikator yang digunakan peneliti relevan secara konseptual untuk menggambarkan pola transaksi berisiko (fraud) dalam konteks penelitian akademik.
 () Setuju (✓) Setuju dengan catatan () Tidak setuju
2. Kelompok fitur yang dinilai relevan (boleh centang lebih dari satu):
 (✓) Nilai & konteks transaksi (*amt, category*)
 (✓) Pola waktu transaksi (*trans_hour, trans_dayofweek, day_period, is_weekend*)
 (✓) Perilaku jangka pendek/velocity (*diff_amt, avg_amt_last_7, count_last_7, category_freq*)
 (✓) Lokasi/jarak sebagai sinyal anomali (*distance_km, state*)
 (✓) Risiko merchant berbasis historis (*merchant_risk_rate*)
 () Lainnya, seperti profil/demografi sebagai proxy (*gender, age, age_group*)

Gambar A.1. Surat Pernyataan Validasi Pakar



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

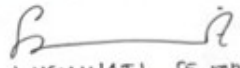
3. Secara keseluruhan, penelitian ini layak digunakan untuk tujuan komparasi metode (XGBoost/LightGBM/CatBoost) dan analisis interpretabilitas (SHAP) pada dataset publik.
(✓) Layak () Layak dengan catatan () Kurang layak

B. Catatan Pakar (opsional, singkat 2-3 poin)

1. Penggunaan fitur age_group terutama pada kelompok usia <18
2. ~~Walaupun~~ ~~apapun~~ diberi catatan karena penggunaan di Indo cenderung sedikit berbeda.
3.

Demikian lembar validasi ini dibuat untuk digunakan sebagai bukti validasi pakar dalam penyusunan Tugas Akhir.

Pekanbaru, 6 / Januari / 2026
Narasumber,


(..... LUSIAWATI, S.E.MBA)
NIP. 19780527 200710 2008

Gambar A.2. Surat Pernyataan Validasi Pakar

LAMPIRAN B DOKUMENTASI

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar B.1. Dokumentasi



DAFTAR RIWAYAT HIDUP



Saya, Amelia Rahmadani, lahir pada 09 November 2003, merupakan mahasiswa Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau. Riwayat pendidikan saya dimulai dari SDN 005 Teluk Binjai Dumai, SMP Negeri 1 Kota Dumai, dan SMAS Budi Dharma Dumai. Selama masa studi di perguruan tinggi, saya aktif mengembangkan kompetensi akademik dan profesional melalui

kegiatan pembelajaran, organisasi, pengembangan diri, serta keterlibatan dalam penelitian dan publikasi ilmiah. Dalam pengalaman profesional, saya telah melaksanakan Kerja Praktik di PT Kilang Pertamina Internasional RU II Dumai serta mengikuti program Merdeka Belajar Kampus Merdeka (MBKM) di PLN Icon Plus, yang memperkuat pemahaman mengenai proses kerja, pengelolaan data, serta kesiapan menghadapi dunia kerja. Pada bidang pengabdian masyarakat, saya melaksanakan Kuliah Kerja Nyata (KKN) di Desa Kadur, Kecamatan Rupat Utara. Saya juga merupakan awardee Beasiswa Bank Indonesia periode 2024–2025 dan aktif dalam komunitas Generasi Baru Indonesia (GenBI). Selain itu, saya pernah mengemban amanah sebagai Sekretaris Umum Himpunan Mahasiswa Sistem Informasi periode 2024, yang melatih kemampuan koordinasi, administrasi, dan komunikasi organisasi. Saya juga aktif sebagai anggota Puzzle Research Data Technology (Predatech) serta terlibat dalam berbagai kepanitiaan dan kegiatan kemahasiswaan. Saya memiliki pengalaman publikasi ilmiah pada jurnal nasional di bidang sistem informasi, tata kelola TI, dan data mining. Ke depan, saya berkomitmen untuk terus meningkatkan kompetensi akademik dan profesional melalui pembelajaran berkelanjutan serta kontribusi dalam kegiatan yang berdampak. Demikian daftar riwayat hidup ini saya susun dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.