TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

Klasifikasi Fraud Pada Transaksi Finansial Melalui Integrasi TabTransformer dan Oversampling Generatif CTGAN

Tangguh Prana Welas Sukma, Ery Hartati*

[1]Fakultas Ilmu Komputer dan Rekayasa, Program Studi Informatika, Universitas Multi Data Palembang, Palembang, Indonesia
Email: [1]ryuokaakagami@gmail.com, [2,*]ery_hartati@mdp.ac.id
Email Penulis Koresponding: ery_hartati@mdp.ac.id

Abstrak–Ketimpangan kelas yang sangat ekstrem pada dataset BankSim (1,2% *fraud*) menjadi hambatan utama dalam membangun sistem deteksi yang andal. Penelitian ini mengusulkan integrasi arsitektur *Deep Learning* TabTransformer dengan teknik *oversampling* generatif *Conditional Tabular GAN* (CTGAN) untuk mengatasi bias kelas mayoritas. Hasil evaluasi kualitas data menunjukkan bahwa CTGAN mampu memproduksi data sintetis dengan skor kualitas keseluruhan mencapai 90,05% dan tingkat korelasi pasangan fitur sebesar 91,63%. Temuan eksperimental membuktikan bahwa model usulan memberikan performa yang paling superior dengan capaian metrik *F1-Score* sebesar 85,34%, *Recall* 81,39%, serta *Balanced Accuracy* mencapai 90,64%. Hasil ini secara signifikan melampaui teknik SMOTE yang mencatatkan *F1-Score* sebesar 83,99% namun mengalami kegagalan dalam kalibrasi probabilitas dengan ambang batas ekstrem sebesar 0,98. Sebaliknya, skenario CTGAN menunjukkan stabilitas ambang batas keputusan yang efisien pada nilai 0,46. Validasi melalui analisis SHAP mengonfirmasi bahwa variabel rekayasa fitur seperti *merchantRisk*, *custStepDiff*, dan *amtZScoreByCat* memberikan kontribusi paling dominan terhadap prediksi model. Penelitian ini menyimpulkan bahwa sinergi paradigma *Data-Centric AI* melalui pendekatan hibrida generatif mampu mewujudkan model klasifikasi yang tangguh, presisi, serta memiliki akuntabilitas tinggi untuk perlindungan sektor perbankan digital.

Kata Kunci: Deteksi Fraud; TabTransformer; CTGAN; Rekayasa Fitur; SHAP

Abstract–Extreme class imbalance in the BankSim dataset (1.2% fraud) is a major hurdle to building reliable detection systems. This study proposes the integration of the TabTransformer architecture with the Conditional Tabular GAN (CTGAN) oversampling technique to address majority class bias. Data quality evaluations indicate that CTGAN produces synthetic data with an overall quality score of 90.05% and a column pair correlation trend of 91.63%. Experimental findings prove the proposed model delivers superior performance, achieving an F1-Score of 85.34%, a Recall of 81.39%, and a Balanced Accuracy of 90.64%. These results significantly outperform the SMOTE technique, which recorded an F1-Score of 83.99% but suffered from probability calibration failure with an extreme optimal threshold of 0.98. In contrast, the CTGAN scenario demonstrates efficient decision threshold stability at 0.46. Validation through SHAP analysis confirms that engineered variables such as merchantRisk, custStepDiff, and amtZScoreByCat provide dominant contributions to model predictions. This research concludes that the synergy of the Data-Centric AI paradigm facilitates the creation of robust, precise, and highly accountable classification models for digital banking protection within financial transaction systems.

Keywords: Fraud Detection; TabTransformer; CTGAN; Feature Engineering; SHAP

1. PENDAHULUAN

Pada era transformasi digital saat ini, lanskap transaksi finansial telah bergeser dari metode konvensional menuju ekosistem digital yang serba cepat. Kemudahan ini memungkinkan aktivitas ekonomi dilakukan tanpa batasan waktu, sehingga mendorong adopsi sistem pembayaran digital secara masif. Berdasarkan Laporan Kebijakan Moneter Triwulan IV 2024 yang dirilis oleh Bank Indonesia (2024), volume transaksi digital sepanjang tahun 2024 mencapai angka fenomenal sebesar 34,5 miliar transaksi, dengan indeks pertumbuhan 36,1% dibandingkan tahun sebelumnya. Data ini mencerminkan ketergantungan besar masyarakat terhadap infrastruktur teknologi finansial saat ini. Namun, di balik efisiensi tersebut, muncul ancaman serius berupa tindak kecurangan transaksi atau *fraud* yang dapat mengganggu stabilitas ekosistem ekonomi.

Secara fundamental, *fraud* merupakan manifestasi kecurangan yang dilakukan oleh entitas tertentu melalui pencurian identitas, transaksi palsu, dan manipulasi data untuk mengelabui protokol keamanan. Dampak dari tindakan ini sangat destruktif, terutama pada kerugian finansial yang harus ditanggung korban maupun lembaga keuangan (Priatna dkk., 2025). Fenomena ini sering dipicu oleh kurangnya kewaspadaan pengguna dalam berinteraksi dengan teknologi, yang membuka celah bagi modus *social engineering* dan penyebaran *malware (Yuhertiana & Amin, 2024)*. Merujuk pada analisis World Bank Group (2023), pola *fraud* terus berevolusi seiring kemajuan teknologi informasi. Kondisi ini terkonfirmasi oleh data Otoritas Jasa Keuangan (OJK) yang mencatat 42.257 laporan penipuan hingga 9 Februari 2025, dengan akumulasi kerugian mencapai Rp700,2 miliar (Putri, 2025). Skala kerugian ini menegaskan bahwa ancaman *fraud* memerlukan solusi deteksi yang cerdas, adaptif, dan berkelanjutan.

Sebagai respons, berbagai institusi memanfaatkan teknologi *Artificial Intelligence* (AI) untuk membangun sistem deteksi cerdas yang mampu mengenali pola kecurangan secara otomatis (Odeyemi dkk., 2024). Penelitian terdahulu oleh Dharmana dkk. (2024) serta Bonde dan Bichanga (2025) telah menunjukkan efektivitas model *Machine Learning* hibrida dalam mengidentifikasi anomali transaksi. Meskipun demikian, masih terdapat ruang pengembangan pada aspek ketajaman deteksi. Karakteristik data *fraud* melibatkan interaksi kompleks antara profil demografis (kategorikal) dan pola transaksional (numerik). Hal ini menuntut model yang mampu memahami relasi kontekstual tabular secara utuh. Studi Padhi dkk. (2021) menunjukkan bahwa arsitektur *Transformer* memiliki potensi besar dalam memodelkan

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

struktur data tabular yang kompleks melalui mekanisme *self-attention*. Mekanisme ini memungkinkan model menangkap hubungan antar-atribut secara dinamis berdasarkan profil entitas yang terlibat.

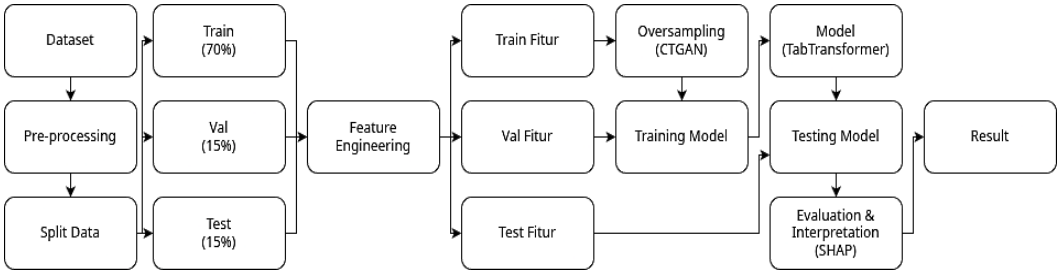
Oleh karena itu, penelitian ini mengusulkan algoritma *TabTransformer (Huang dkk., 2020)* untuk klasifikasi *fraud*. Keunggulan utamanya terletak pada penggunaan mekanisme *self-attention* untuk mengekstraksi hubungan kontekstual mendalam antar-fitur kategorikal yang sering terabaikan oleh model konvensional. Efektivitas arsitektur ini diperkuat oleh penelitian Conteh dan Zhou (2025) yang mencatat *F1-Score* sebesar 99,94% melalui pendekatan *Cost-Sensitive Learning*. Namun, tantangan mendasar dalam klasifikasi *fraud* adalah ketidakseimbangan data (*class imbalance*), di mana sampel kecurangan jauh lebih sedikit dibandingkan transaksi normal (Assabil, 2024). Fenomena ini termanifestasi nyata pada dataset BankSim (Lopez-Rojas & Axelsson, 2014) yang merepresentasikan distribusi ekstrem dunia nyata, dengan transaksi *fraud* hanya berjumlah 1,2%. Jika tidak ditangani, model cenderung bias terhadap kelas mayoritas dan gagal mengenali pola kecurangan secara akurat (Ibrahim & Alfauzan, 2025).

Guna menanggulangi kendala tersebut, strategi penyeimbangan distribusi kelas melalui *oversampling* menjadi krusial. Dalam berbagai literatur, *Synthetic Minority Over-sampling Technique* (SMOTE) sering menjadi rujukan melalui mekanisme interpolasi linear antar-sampel (Chawla dkk., 2002). Kendati demikian, SMOTE memiliki limitasi karena pembentukan data sintetisnya hanya terbatas pada garis lurus antar-titik, yang berpotensi memicu *noise* dan kegagalan dalam merepresentasikan variabilitas data kompleks. Sebagai solusi inovatif, Xu dkk. (2019) memperkenalkan *Conditional Tabular GAN* (CTGAN) berbasis *Deep Learning*. Keunggulan CTGAN terletak pada kapasitasnya mengekstraksi distribusi probabilitas data secara utuh melalui prinsip *Generative Adversarial Networks (Goodfellow dkk., 2016)*. Pendekatan ini divalidasi oleh Nugraha dkk. (2022) dengan performa di atas 90% dibanding SMOTE. Temuan ini selaras dengan studi Alshawi (2023) dan Patil (2021) yang menegaskan keunggulan CTGAN dalam memodelkan distribusi data rumit untuk klasifikasi.

Kesenjangan penelitian (*research gap*) dalam studi ini terletak pada masih jarangnya integrasi antara pendekatan generatif modern CTGAN dengan arsitektur berbasis *attention* seperti TabTransformer pada dataset dengan ketimpangan ekstrem. Mayoritas penelitian sebelumnya, seperti Bonde dan Bichanga (2025), masih berfokus pada kombinasi SMOTE dengan algoritma *ensemble tree* seperti *Random Forest* atau *XGBoost*. Meskipun efektif, model *ensemble tree* memiliki keterbatasan dalam menangkap hubungan semantik fitur kategorikal dibandingkan mekanisme *attention*. Secara spesifik, urgensi penerapan strategi hibrida ini didorong oleh kompleksitas topologi data *fraud* yang sering kali menempati ruang *manifold non-linear* yang sulit dijangkau oleh interpolasi sederhana. Dalam konteks ini, CTGAN bertindak lebih dari sekadar penambah jumlah sampel, ia merekonstruksi korelasi antar-variabel yang valid secara statistik, sehingga mencegah distorsi informasi yang kerap terjadi pada metode klasik. Data sintetis berkualitas tinggi tersebut kemudian dioptimalkan oleh mekanisme *Contextual Embeddings* pada TabTransformer, yang mampu mengubah fitur kategorikal berdimensi tinggi menjadi representasi vektor yang padat dan bermakna. Sinergi antara fidelitas data dari generator dan kapasitas representasi dari model berbasis *attention* ini dihipotesiskan mampu memecahkan masalah *vanishing gradient* pada data minoritas yang sering dialami oleh *Deep Learning* standar. Kontribusi ilmiah penelitian ini adalah menghadirkan pendekatan hibrida *Deep Generative* dan *Self-Attention* untuk mengisi celah tersebut. Pendekatan ini diharapkan menghasilkan model deteksi yang lebih *robust* dan akurat dalam mengenali pola kecurangan di tengah kondisi data yang tidak seimbang, sekaligus menekan risiko *False Negative* yang menjadi parameter paling kritis dalam mitigasi kerugian finansial perbankan. Selain itu, validasi melalui pendekatan *Explainable AI* (XAI) (Lundberg & Lee, 2017) turut diintegrasikan guna menjamin bahwa peningkatan performa tersebut tetap mematuhi prinsip transparansi algoritma yang diwajibkan dalam regulasi industri keuangan, menjadikan usulan ini sebagai solusi yang komprehensif baik dari sisi akurasi teknis maupun akuntabilitas operasional.

2. METODOLOGI PENELITIAN

Penelitian eksperimental ini mengklasifikasi *fraud* melalui tahapan sistematis yang meliputi *data collection*, *pre-processing*, *splitting*, *feature engineering*, *oversampling*, *training model*, hingga *model evaluation & interpretation* (SHAP), sebagaimana diilustrasikan pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

Pemilihan algoritma *TabTransformer* didasarkan pada kemampuannya dalam menangkap hubungan kontekstual pada data tabular yang sering kali terabaikan oleh model *tree-based* konvensional maupun *Deep Learning* standar (MLP). Sementara itu, CTGAN dipilih karena keunggulannya dibandingkan dengan SMOTE dalam memodelkan

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

distribusi probabilitas yang kompleks. Penelitian ini menghipotesiskan bahwa integrasi pendekatan *generatif* (CTGAN) dengan arsitektur berbasis atensi (*TabTransformer*) akan menghasilkan model deteksi *fraud* yang tidak hanya memiliki sensitivitas tinggi (*Recall*) terhadap kelas minoritas, tetapi juga menjaga presisi (*Precision*) dengan menekan angka *false positive* dibandingkan metode interpolasi tradisional.

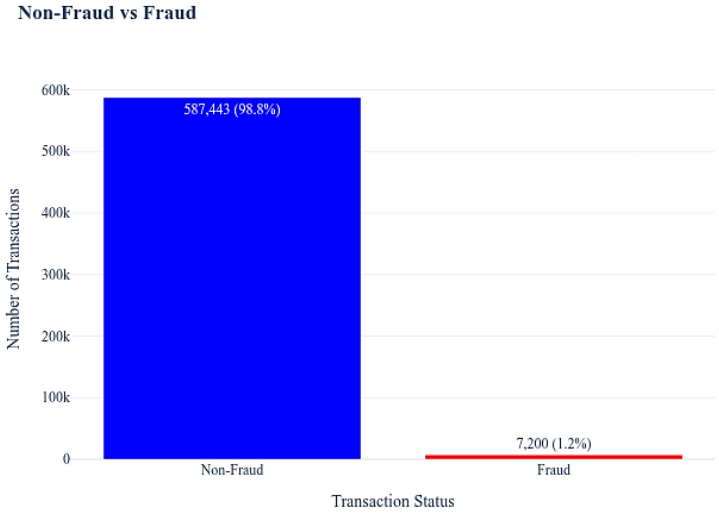
2.1 *Data Collection*

Penelitian kali ini menggunakan dataset BankSim (Lopez-Rojas & Axelsson, 2014), yang terdiri dari 594.643 log transaksi hasil simulasi bank di Spanyol. Penggunaan data sintetis ini dipilih sebagai alternatif standar mengingat batasan privasi ketat pada data finansial riil. Dataset ini memuat 9 fitur (7 kategorikal dan 2 numerik) dengan spesifikasi lengkap yang dirangkum pada Tabel 1.

Tabel 1. Spesifikasi Variabel pada Dataset

Fitur	Tipe Data	Jumlah Unik	Keterangan
step	*int64*	180	Langkah Waktu Simulasi
customer	*object*	4.112	ID Customer
age	*object*	8	Kategori Usia
gender	*object*	4	Jenis Kelamin
zipCodeOri	*object*	1	Kode Pos Customer
merchant	*object*	50	ID Merchant
zipMerchant	*object*	1	Kode Pos Merchant
category	*object*	15	Kategori Merchant
amount	*float64*	23.767	Nominal Uang
fraud	*int64*	2	Target

Tantangan utama dari dataset ini adalah ketidakseimbangan kelas (*imbalanced data*) yang ekstrem, di mana proporsi *fraud* hanya sebesar 1,2% dibandingkan 98,8% transaksi normal. Untuk visualisasi kelas tersebut disajikan pada Gambar 2.



Gambar 2. Distribusi Kelas Target Pada Fitur Fraud

Berdasarkan visualisasi distribusi tersebut, terlihat jelas dominasi absolut kelas mayoritas. Ketimpangan ekstrem ini mengindikasikan bahwa tanpa penanganan khusus, model akan memiliki bias yang kuat terhadap transaksi normal dan berpotensi gagal mengenali pola *fraud.*

2.2 *Pre-processing & Initial Features*

Sebagai referensi kondisi awal, visualisasi sampel data mentah (raw data) sebelum pemrosesan disajikan pada Gambar 3.

	step	customer	age	gender	zipcodeOri	merchant	zipMerchant	category	amount	fraud
0	0	'C1093826151'	'4'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	4.55	0
1	0	'C352968107'	'2'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	39.68	0
2	0	'C2054744914'	'4'	'F'	'28007'	'M1823072687'	'28007'	'es_transportation'	26.89	0
3	0	'C1760612790'	'3'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	17.25	0
4	0	'C757503768'	'5'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	35.72	0

Gambar 3. Sampel *Raw* Data BankSim

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

Berdasarkan analisis statistik pada data tersebut, proses pembersihan dilakukan dengan membuang fitur *zipCodeOri* dan *zipCodeMerchant* yang memiliki variansi nol (*cardinality* = 1), diikuti standarisasi format fitur kategorikal. Selanjutnya, fitur temporal *custStepDiff* dan *merchStepDiff* diinisialisasi guna menangkap dinamika jeda waktu antar transaksi. Nilai ini dihitung sebagai selisih *step* saat ini dengan *step* sebelumnya untuk entitas yang sama, mengacu pada Persamaan (1).

$$\Delta t_i = Step_{(i,t)} - Step(i, t\text{-}1) \quad (1)$$

Operasi ini secara alami menghasilkan nilai kosong (*NaN*) pada transaksi pertama entitas karena ketiadaan referensi waktu sebelumnya, sebagaimana diilustrasikan pada Tabel 2. Nilai *NaN* ini sengaja dipertahankan sementara dan baru ditangani melalui imputasi pasca-*splitting* untuk mencegah kebocoran data (*data leakage*).

Tabel 2. Representasi Data Pasca-Inisialisasi Fitur Waktu

step	customer	custStepDiff	merchant	merchStepDiff
0	C1093826151	NaN	M348934600	NaN
0	C352968107	NaN	M348934600	0.0
0	C2054744914	NaN	M1823072687	NaN
0	C1760612790	NaN	M348934600	0.0
0	C757503768	NaN	M348934600	0.0

2.3 *Splitting Data*

Setelah melalui tahap pre-processing, splitting data dilakukan dalam dua tahap menggunakan teknik stratified shuffle split untuk menghasilkan *subset train data*, *val data*, dan *test data* dengan rasio 70:15:15. Penggunaan teknik ini menjamin bahwa setiap data merepresentasikan distribusi kelas asli, terutama kelas fraud yang bersifat minoritas. Rincian mengenai proporsi jumlah data untuk setiap *subset*, baik kelas *non-fraud* maupun *fraud*, dipaparkan secara terperinci pada Tabel 3.

Tabel 3. Proporsi Kelas Target pada Setiap Subset Data

Data	Jumlah Data	Jumlah Fraud	Rasio Persentase Fraud
Train	416.250	5.040	1.21%
Val	89.197	1.080	1.21%
Test	89.196	1.080	1.21%

Hasil pembagian tersebut mengonfirmasi bahwa teknik *stratified shuffle split* berhasil mempertahankan konsistensi rasio *fraud* sebesar 1,21% di seluruh *subset*. Konsistensi ini krusial untuk menjamin validitas evaluasi model agar tidak bias akibat pergeseran distribusi data.

2.4 *Feature Engineering*

Pada tahapan ini, data latih (*training set*) digunakan untuk merekayasa fitur guna mendapatkan representasi yang lebih mendalam dari data mentah. Parameter hasil perhitungan dari data latih kemudian diterapkan pada data validasi dan data uji. Pendekatan ini diterapkan secara ketat untuk mencegah kebocoran data (*data leakage*) yang dapat menyebabkan bias pada evaluasi model. Proses ini menghasilkan 13 fitur baru yang dikelompokkan sebagai berikut:

2.5 *Temporal Imputation*

Pada tahapan sebelumnya, teridentifikasi bahwa fitur *custStepDiff* dan *merchStepDiff* memiliki nilai kosong (*NaN*) pada setiap transaksi pertama. Oleh karena itu, penanganan dilakukan dengan teknik imputasi menggunakan nilai median global dari data latih (*training set*). Strategi imputasi median dipilih karena lebih resisten terhadap pencilan (*outliers*) dibandingkan rata-rata (*mean*). Hal ini dikonfirmasi oleh statistik data latih yang menunjukkan *skewness* signifikan pada fitur selisih waktu (*custStepDiff Mean*: 1.21 vs *Median*: 1.00; *merchStepDiff Mean*: 0.01 vs *Median*: 0.00). Penggunaan median (1.00 dan 0.00) memastikan validitas data tetap terjaga tanpa distorsi nilai ekstrem.

2.5.1 *Frequency Encoding*

Pada tahapan ini, frekuensi kemunculan entitas (*customer* dan *merchant*) pada data latih dihitung dan dipetakan menjadi fitur *custFreq* dan *merchFreq*. Untuk menangani data baru (*unseen data*) yang tidak memiliki riwayat, diterapkan mekanisme imputasi menggunakan nilai median global guna menghindari bias akibat distribusi *long-tail*, sebagaimana diformulasikan pada Persamaan (2).

$$Freq(x) = \begin{cases} Count(x, D_{train}) \text{ jika } x \in D_{train} \\ Median(F_{train}) \text{ jika } x \notin D_{train} \end{cases} \quad (2)$$

Pemilihan median dinilai lebih *robust* dibandingkan rata-rata (*mean*) karena teridentifikasi adanya kesenjangan ekstrem pada distribusi *merchant* (Mean: 8.325 vs Median: 416). Disparitas ini menunjukkan bahwa rata-rata terdistorsi

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

oleh segelintir *merchant* besar, sehingga penggunaan median memberikan estimasi yang lebih stabil untuk entitas yang belum dikenali.

2.5.2 *Velocity*

Pada tahapan ini, fitur *custVelocity* dan *merchVelocity* dikonstruksi untuk menangkap pola kecepatan pengurasan dana. Nilai ini dihitung sebagai rasio nominal terhadap jeda waktu (Δt_i), dengan penambahan konstanta 1 pada penyebut guna mencegah pembagian dengan nol, sebagaimana Persamaan (3).

$$Velocity_{i,t}=\frac{Amount_t}{\Delta t_i+1} \tag{3}$$

Analisis distribusi menunjukkan bahwa median *velocity* transaksi *fraud* jauh melampaui transaksi normal (Customer: 139,10 vs 12,85; Merchant: 285,26 vs 26,46). Berdasarkan disparitas ini, fitur *distCustVeloFraud* dan *distMerchVeloFraud* dibuat untuk mengukur jarak absolut *velocity* saat ini terhadap profil median *fraud*, mengikuti Persamaan (4).

$$Dist_{i,t}=\left|Velocity_{i,t}-\tilde{V}_{fraud}\right| \tag{4}$$

2.5.3 *ZScore & Distance*

Pada tahapan ini, fitur *amtZScoreByCat* dikonstruksi menggunakan teknik *Z-Score* per kategori (*category-wise standardization*) untuk mengukur tingkat ekstremitas penyimpangan nominal transaksi dari rata-rata kategori belanjanya. Konstanta $\epsilon=10^{-6}$ ditambahkan pada penyebut guna menjamin stabilitas numerik, sebagaimana diformulasikan pada Persamaan (5).

$$Z_t=\frac{Amount_t-\mu_{cat}}{\sigma_{cat}+\epsilon} \tag{5}$$

Selanjutnya, berdasarkan temuan bahwa median transaksi *fraud* jauh lebih tinggi dibandingkan normal (317,92 vs 26,58), dibuat fitur *distAmtFraudMed*. Fitur ini mengukur jarak absolut antara nominal transaksi saat ini dengan profil median *fraud* untuk menangkap anomali nilai tinggi, mengikuti Persamaan (6).

$$Dist_t=\left|Amount_t-\tilde{A}_{fraud}\right| \tag{6}$$

2.5.4 *Risk Scoring*

Pada tahapan ini, variabel kategorikal (*merchant*, *category*, dan *cohort*) dikonversi menjadi skor risiko numerik (*merchantRisk*, *categoryRisk*, *cohortRisk*) menggunakan teknik *Target Encoding*. Pendekatan ini menghitung probabilitas terjadinya *fraud* (rata-rata target) untuk setiap entitas. Guna memitigasi *overfitting* pada kategori dengan frekuensi rendah (*rare labels*), diterapkan mekanisme *smoothing* dengan parameter α=10. Teknik ini menyeimbangkan statistik lokal kategori dengan rata-rata global untuk meredam *noise*, sebagaimana diformulasikan pada Persamaan (7).

$$isk(x)=\frac{(n\cdot\mu_{local})+(\alpha\cdot\mu_{global})}{n+\alpha} \tag{7}$$

Dalam formulasi tersebut, n merepresentasikan jumlah transaksi pada kategori terkait, sedangkan μ_{local} dan μ_{global} masing-masing menyatakan rata-rata target (probabilitas fraud) pada tingkat kategori dan keseluruhan data. Mekanisme ini juga menjamin stabilitas model terhadap data baru (*unseen data*), di mana kategori yang tidak dikenali otomatis diisi dengan nilai global.

2.6 *Oversampling*

Setelah melalui proses *feature engineering*, dilakukan tahapan *oversampling* data untuk menangani ketidakseimbangan data, dimana pada penelitian ini diterapkan metode *oversampling* menggunakan algoritma CTGAN (*Conditional Tabular GAN*). Proses ini tidak sekadar menduplikasi data, melainkan mensintesis data baru berdasarkan distribusi probabilitas bersyarat dari kelas minoritas (*fraud*).

Mekanisme pembentukan data sintetis diawali dengan menghitung probabilitas setiap baris (*row*) berdasarkan kondisi kolom kategorikal tertentu (D_{i^*}). Dalam penelitian ini, kondisi dikunci pada label *fraud*, di mana probabilitasnya dihitung mengikuti persamaan yang diformulasikan pada Persamaan (8).

$$P(\text{row})=\sum_{k\in D_{i^*}} P_G\left(\text{row}|D_{i^*}=k^*\right)P\left(D_{i^*}=k\right) \tag{8}$$

Berdasarkan probabilitas tersebut, generator kemudian melakukan pengambilan sampel (*sampling*) secara spesifik pada kategori target (k^*) untuk menghasilkan data sintetis $(\tilde{r})$ yang memiliki karakteristik *fraud*. Proses sampling ini diformulasikan dalam Persamaan (9).

$$\tilde{r} \sim P_G\left(\text{row}|D_{i^*}=k^*\right) \tag{9}$$

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

Tahap pelatihan sistem ini melibatkan optimasi melalui skema persaingan *min-max* antara *generator* g dan *discriminator* d. Dalam mekanisme tersebut, *generator* dipacu untuk memproduksi data yang mampu meminimalkan akurasi deteksi *discriminator*, sementara *discriminator* dilatih secara intensif guna memaksimalkan kemampuannya dalam menolak setiap sampel sintetis. Adapun formulasi dari fungsi objektif yang mendasari proses optimisasi tersebut dijabarkan secara matematis melalui Persamaan (10).

$$\min_{g}\max_{d} V(g,d)=E_{x\sim P_{data}}[\log d(x)]+E_{z\sim P_z}\left[\log\left(1-d\left(g(z)\right)\right)\right] \quad (10)$$

Tahapan implementasi diawali dengan penyusunan skema metadata guna memisahkan antara atribut kategorikal dan numerik, yang dilanjutkan dengan proses pelatihan model selama 1.000 epoch. Rincian konfigurasi *hyperparameter* yang diterapkan dirangkum pada Tabel 4.

Tabel 4. Konfigurasi *Hyperparameter* CTGAN

Parameter	Nilai	Keterangan
Epochs	1.000	Jumlah iterasi pelatihan penuh
Batch Size	500	Jumlah sampel per *update* gradien
PAC	10	Jumlah sampel per paket diskriminator
Embedding Dim	128	Dimensi representasi fitur diskrit
Generator Dim	(256, 256)	Dimensi *hidden layer* generator
Discriminator Dim	(256, 256)	Dimensi *hidden layer* diskriminator
Learning Rate	2×10^{-4}	Laju pembelajaran (*G* dan *D*)

Penerapan parameter tersebut, khususnya jumlah *epoch* yang tinggi (1.000) dan dimensi *embedding* (128), ditujukan untuk memastikan *generator* memiliki kapasitas yang memadai dalam mempelajari distribusi data yang kompleks, sekaligus menjaga stabilitas *gradien* selama proses sintesis berlangsung.

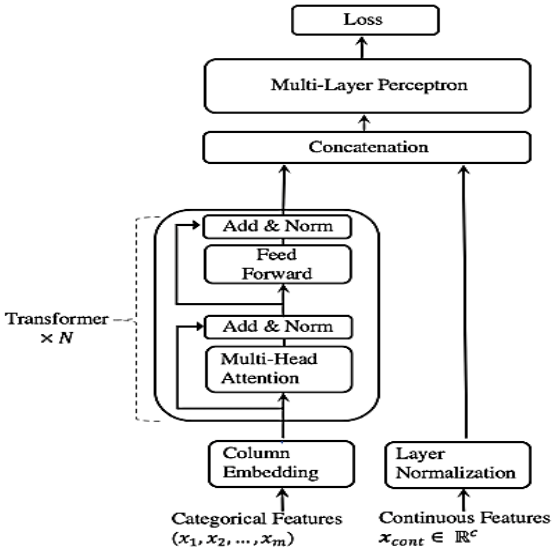
2.7 *Training Model*

Pada tahapan ini, dilakukan pelatihan model klasifikasi menggunakan algoritma *TabTransformer* (Huang dkk., 2020). Berbeda dengan MLP standar yang memperlakukan semua fitur secara independen, arsitektur ini dipilih karena keunggulannya dalam memodelkan interaksi antar fitur kategorikal secara mendalam melalui mekanisme *attention.* Secara operasional, fitur kategorikal ditransformasi menjadi *dense vectors* melalui lapisan *embedding*, sedangkan fitur numerikal distandarisasi menggunakan *Layer Normalization*.

Inti kekuatan model ini terletak pada mekanisme *Multi-Head Self-Attention* di jalur kategorikal yang bertugas menangkap ketergantungan kontekstual antar fitur. Mekanisme ini menghitung bobot secara dinamis mengikuti formulasi persamaan (11).

$$Attention(Q,K,V)=softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (11)$$

Output dari lapisan *transformer* kemudian digabungkan (*concatenated*) dengan fitur numerikal untuk menghasilkan prediksi akhir melalui lapisan *Multi-Layer Perceptron* (MLP), sebagaimana diilustrasikan pada Gambar 4.



Gambar 4. Representasi Struktural Model *TabTransformer*

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

Skema arsitektur tersebut memperlihatkan mekanisme pemrosesan hibrida, di mana fitur kategorikal mendapatkan perhatian khusus melalui self-attention untuk menangkap konteks semantik, sebelum disatukan dengan fitur numerik untuk klasifikasi akhir.

Adapun Proses pelatihan dijalankan menggunakan fungsi kerugian Binary Cross Entropy with Logits dan dioptimasi menggunakan algoritma AdamW dengan mekanisme *learning rate scheduler* (ReduceLROnPlateau). Guna memaksimalkan performa pada kelas minoritas, ambang batas keputusan (*threshold*) dievaluasi secara dinamis berdasarkan *F1-Score* terbaik pada data validasi. Rincian konfigurasi *hyperparameter* yang digunakan dirangkum dalam Tabel 5.

Tabel 5. Konfigurasi *Hyperparameter* Algoritma *TabTransformer*

Parameter	Nilai	Keterangan
Embedding Dim	32	Dimensi vektor fitur kategorikal
Transformer Depth	6	Jumlah lapisan *encoder transformer*
Attention Heads	8	Jumlah *head* pada mekanisme atensi
Attention Dropout	0.1	Probabilitas *dropout* pada lapisan atensi
FF Dropout	0.1	Probabilitas *dropout* pada *feed-forward*
MLP Hidden Layers	-128,64	Dimensi lapisan tersembunyi (hasil *mults* 4 & 2)
MLP Activation	ReLU	Fungsi aktivasi pada lapisan MLP
Output Dim	1	Dimensi output (Klasifikasi Biner)
Batch Size	128	Ukuran sampel per iterasi
Learning Rate	1×10^{-3}	Laju pembelajaran awal (AdamW)
Weight Decay	1×10^{-4}	Regularisasi bobot optimizer
Epochs	25	Total durasi pelatihan
Loss Function	BCEWithLogitsLoss	Fungsi kerugian entropi biner
LR Scheduler	ReduceLROnPlateau	Penyesuaian *learning rate* berdasarkan performa validasi

Kombinasi parameter tersebut, khususnya penggunaan *weight decay* dan *dropout*, digunakan secara spesifik untuk mencegah *overfitting*, sementara *scheduler* adaptif memastikan konvergensi model yang stabil pada *loss surface* yang kompleks.

2.8 *Model Evaluation & Interpretation* (SHAP)

Evaluasi metode dilakukan secara komprehensif mencakup kinerja prediktif dan interpretabilitas model menggunakan *SHapley Additive exPlanations* (SHAP). Pendekatan ganda ini diterapkan untuk memastikan model *TabTransformer* tidak hanya unggul secara statistik, tetapi juga memenuhi standar akuntabilitas perbankan.

2.8.1 *Comparative Experimental Scenarios*

Untuk mengukur dampak pendekatan *oversampling* secara objektif, penelitian ini merancang tiga skenario eksperimen. Seluruh skenario divalidasi menggunakan data uji (*test set*) asli yang tidak dimodifikasi untuk mencegah bias evaluasi.

1. *Baseline Scenario* (*Imbalanced Data*): Model dilatih menggunakan data asli tanpa modifikasi distribusi sebagai kontrol batas bawah (*lower bound*) untuk mengukur kegagalan deteksi pada kondisi ketimpangan ekstrem.
2. *Comparative Scenario* (SMOTE): Model dilatih menggunakan data yang diseimbangkan dengan teknik SMOTE sebagai pembanding standar industri.
3. *Proposed Scenario* (CTGAN): Model dilatih menggunakan data campuran sintetis CTGAN untuk menguji hipotesis bahwa model generatif mampu menangkap distribusi probabilitas gabungan (*joint distribution*) lebih baik dibandingkan interpolasi linear.

Perlu ditekankan bahwa seluruh skenario di atas diuji menggunakan data uji (*test set*) yang tidak dimodifikasi oleh teknik *oversampling*. Langkah ini diterapkan secara ketat untuk mencegah bias evaluasi, memastikan model diuji pada kondisi distribusi alami (*real-world distribution*) dimana proporsi *fraud* tetap minoritas.

2.8.2 *Threshold Optimization & Evaluation Metrics*

Tahap evaluasi diawali dengan memulihkan bobot model paling optimal (*best checkpoint*) hasil fase validasi guna mengantisipasi efek *overfitting* pasca pelatihan. Dalam menentukan klasifikasi akhir, penggunaan ambang batas standar sebesar 0,5 dipandang kurang efektif untuk menangani data dengan ketimpangan kelas yang ekstrem. Sebagai solusinya, diterapkan mekanisme *Threshold Tuning* guna mengidentifikasi titik batas probabilitas yang paling dinamis. Pendekatan ini berfokus pada maksimisasi metrik *F1-Score*, yang secara esensial menyatukan nilai *precision* dan *recall* ke dalam satu keseimbangan numerik. Melalui strategi ini, sistem mampu menekan risiko transaksi kecurangan yang tidak terdeteksi (*False Negative*) dengan tetap menjaga stabilitas alarm palsu. Klasifikasi akhir kemudian ditentukan berdasarkan apakah probabilitas sampel melampaui ambang batas optimal yang telah ditetapkan tersebut.

Sebagai konsekuensi dari pendekatan ini, akurasi global tidak dijadikan acuan utama. Evaluasi difokuskan pada metrik yang sensitif terhadap kelas minoritas, yaitu *Precision*, *Recall*, *F1-Score*, dan *Balanced Accuracy*, serta analisis

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

kurva ROC (*Receiver Operating Characteristic*) untuk mengukur kemampuan diskriminatif model secara menyeluruh. Formulasi metrik evaluasi dapat dilihat pada persamaan (12), (13), (14), (15), dan (16).

$$\frac{TP}{TP+FP} \tag{12}$$

$$\frac{\text{TP}}{\text{TP+FN}} \tag{13}$$

$$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{14}$$

$$\frac{\text{Sensitivity+Specificity}}{2} \tag{15}$$

$$\int_0^1 \text{TPR(fpr)d(fpr)} \tag{16}$$

2.8.3 *Shapley Additive exPlanations* (SHAP)

Meskipun model *Deep Learning* seperti TabTransformer memiliki kinerja tinggi, arsitekturnya yang kompleks sering dikategorikan sebagai "kotak hitam" (*black-box*). Untuk mengatasi kendala interpretabilitas ini, diterapkan metode *SHapley Additive exPlanations* (SHAP) sebagaimana diperkenalkan oleh Lundberg dan Lee (2017). Metode ini mengadopsi teori permainan kooperatif (*game theory*) untuk menghitung kontribusi marginal setiap fitur terhadap hasil prediksi akhir.

Dalam penelitian ini, analisis SHAP memiliki peran ganda, sebagai alat interpretasi keputusan model dan sebagai metode validasi fitur. Analisis dilakukan dalam dua tingkatan.

1. *Feature Importance*: Analisis ini digunakan untuk mengidentifikasi hierarki fitur yang paling dominan dalam mendeteksi pola *fraud* secara keseluruhan. Fitur dengan nilai rata-rata SHAP absolut tertinggi dianggap memiliki pengaruh terbesar. Secara khusus, analisis ini bertujuan untuk memvalidasi efektivitas tahapan *Feature Engineering*. Jika fitur-fitur turunan (seperti *velocity*, *risk scores*, dan *z-score*) menduduki peringkat atas dalam *summary plot*, hal tersebut menjadi bukti empiris bahwa rekayasa fitur berhasil mengekstraksi informasi krusial yang tidak dapat ditangkap oleh fitur mentah (*raw features*).
2. *Instance-level Explanation*: Analisis ini digunakan untuk membedah alasan keputusan model pada level transaksi individu. Pendekatan ini memungkinkan analisis *why* (mengapa) sebuah transaksi spesifik ditandai sebagai *fraud*, dengan memvisualisasikan bagaimana kombinasi nilai fitur tertentu mendorong probabilitas prediksi ke arah positif atau negatif.

Penerapan SHAP ini memastikan bahwa model mempelajari pola logis yang sesuai dengan karakteristik domain finansial, serta membuktikan bahwa peningkatan kinerja model didorong oleh kualitas fitur yang dikonstruksi, bukan sekadar menghafal *noise* pada data latih.

3. HASIL DAN PEMBAHASAN

3.1 *Synthetic Data Quality Analysis* (CTGAN)

Evaluasi kualitas data sintetis dilakukan menggunakan kerangka kerja *Synthetic Data Vault* (SDV) untuk menjamin kemiripan karakteristik statistik antara data hasil generasi dan data asli. Ringkasan hasil evaluasi tersebut disajikan pada Tabel 6.

Tabel 6. Skor Evaluasi Kualitas Data

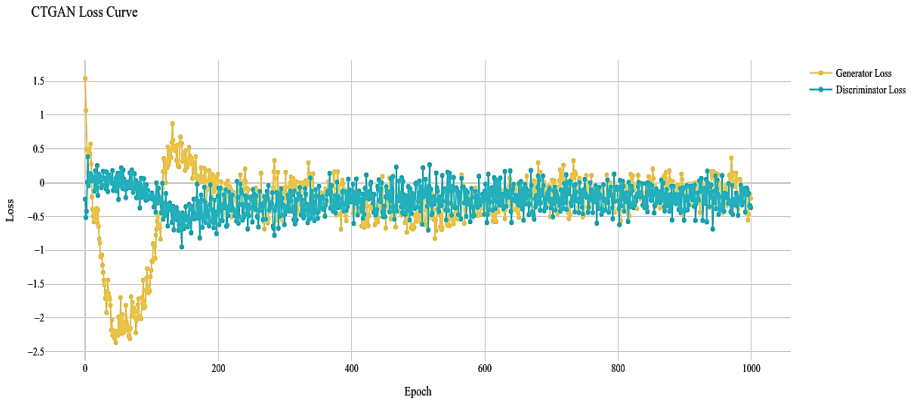
Metrik Evaluasi	Skor	Keterangan
Data Validity	100%	Kepatuhan data terhadap batasan nilai (*min*/*max*) dan tipe data.
Data Structure	100%	Integritas struktur tabel dan kelengkapan kolom.
Column Shapes	89,46%	Kemiripan distribusi marginal per fitur (*univariate distribution*).
Columns Pair Trends	91,63%	Kemiripan pola korelasi antar pasangan fitur (*bivariate correlation*).
Overall Quality Score	90,05%	Rata-rata kualitas data sintetis secara menyeluruh.

Evaluasi metrik menunjukkan kinerja model yang optimal dengan *overall score* mencapai 90,05%. Skor sempurna tercatat pada *validity* dan *structure data*, yang mampu menjamin integritas tipe data dan konsistensi skema tabel. Secara spesifik, model berhasil mereplikasi fitur target *fraud* dengan skor 1.0, serta menangkap pola fitur yang kompleks dari hasil rekayasa seperti *custStepDiff* dan *distMerchVeloFraud* dengan skor tinggi berkisar 96-97%. Meskipun terdapat sedikit variasi pada ekor distribusi fitur *merchantRisk* dengan skor 82%, dalam hal ini pola korelasi antar variabel tetap terjaga dengan baik ditunjukkan pada skor *pair trends* sebesar 91,63%. Secara khusus, skor *Column Shapes* sebesar 89,46%—meskipun merupakan metrik terendah dalam evaluasi ini—masih berada jauh di atas ambang batas toleransi reliabilitas data sintetis. Penurunan minor pada metrik ini merupakan konsekuensi alami dari kompleksitas distribusi *long-tail* pada data keuangan, di mana model generatif harus bekerja ekstra keras untuk menyeimbangkan antara mereplikasi frekuensi transaksi nominal kecil yang dominan dengan transaksi nominal besar

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

yang sangat jarang. Namun, fakta bahwa skor ini tetap mendekati 90% menegaskan bahwa CTGAN mampu menangani tantangan *statistical skewness* jauh lebih baik dibandingkan metode interpolasi tradisional.

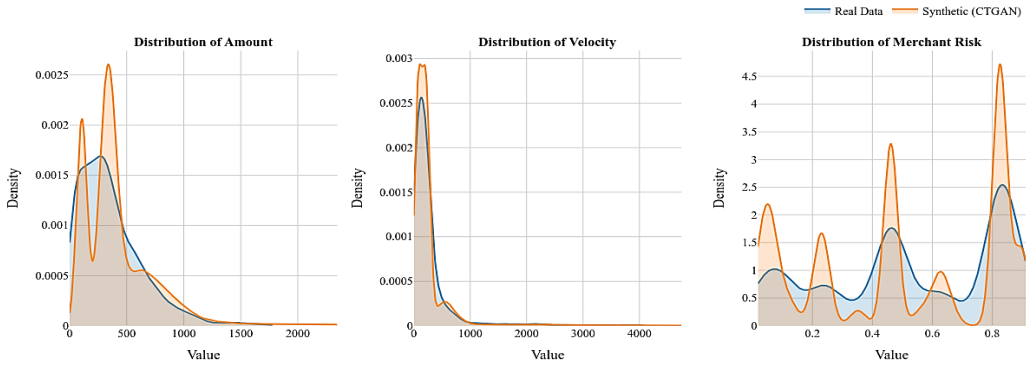
Validitas skor kualitas data di atas tidak terlepas dari stabilitas proses pelatihan model generatif itu sendiri. Untuk memastikan bahwa CTGAN tidak mengalami kegagalan pelatihan (*mode collapse*) atau divergensi, dinamika pembelajaran dipantau melalui pergerakan fungsi kerugian (*loss function*). Visualisasi riwayat pelatihan *Generator* dan *Discriminator* selama 1.000 *epoch* dapat dilihat pada Gambar 5.



Gambar 5. *Training History* CTGAN

Dinamika pembelajaran yang tervisualisasi memperlihatkan kompetisi (*min-max game*) yang sehat antara *Generator* (garis kuning) dan *Discriminator* (garis tosca). Pada fase awal (0-200 epoch), terjadi fluktuasi tajam yang menandakan fase eksplorasi ruang vektor, di mana *Discriminator* masih dengan mudah membedakan data palsu (ditandai dengan *loss* yang rendah pada garis *tosca*). Namun, memasuki pertengahan hingga akhir pelatihan, kedua kurva mulai mencapai titik keseimbangan (*equilibrium*). Pola osilasi yang stabil dan saling berhimpit di sekitar titik nol pada 400 epoch terakhir mengindikasikan bahwa model telah mencapai konvergensi optimal. Hal ini berarti *Generator* berhasil memproduksi data sintetis yang cukup realistis sehingga menyulitkan *Discriminator* untuk membedakannya dari data asli.

Untuk memvalidasi statistik tersebut, diperlukan visualisasi grafik yang menampilkan estimasi densitas *kernel* (*Kernel Density Estimation/KDE*) yang digunakan pada 3 fitur representatif guna memverifikasi apakah model mampu menangkap bentuk distribusi data yang variasi. Adapun visualisasinya dapat dilihat pada Gambar 6.



Gambar 6. Perbandingan Distribusi *Fitur Amount*, *Velocity*, dan *Merchant Risk* (Data Riil vs Sintetis)

Visualisasi tersebut memperlihatkan keselarasan topologi yang signifikan antara data sintetis (CTGAN) dan data asli. Kurva densitas (*density curve*) menunjukkan kemampuan adaptasi model yang presisi dalam mereplikasi berbagai bentuk distribusi. Adapun analisis mendalam terhadap karakteristik distribusi pada ketiga fitur representatif tersebut dipaparkan dalam rincian di bawah ini.

1. *Distribution Of Amount*: Fitur ini memiliki karakteristik distribusi yang sangat condong (*highly skewed*). Model terbukti berhasil menangkap pola global ini, di mana densitas tertinggi terkonsentrasi pada nilai rendah dan melandai secara gradual ke arah nominal besar membentuk ekor panjang (*long-tail*). Konsistensi ini membuktikan bahwa model mampu mereplikasi karakteristik nominal transaksi yang wajar sekaligus mempertahankan probabilitas kemunculan nilai transaksi besar yang jarang terjadi (*rare events*).
2. *Distribution Of Velocity*: Pada fitur ini, terlihat adanya superposisi (himpitan) kurva yang hampir sempurna antara data sintetis dan data asli. Fenomena visual ini mengonfirmasi validitas skor statistik *Column Shapes* yang tinggi (97%) pada evaluasi sebelumnya. Hal ini mengindikasikan bahwa model tidak hanya mempelajari data mentah, tetapi juga mampu menangkap pola fitur turunan yang kompleks dengan akurasi tinggi tanpa distorsi yang berarti.

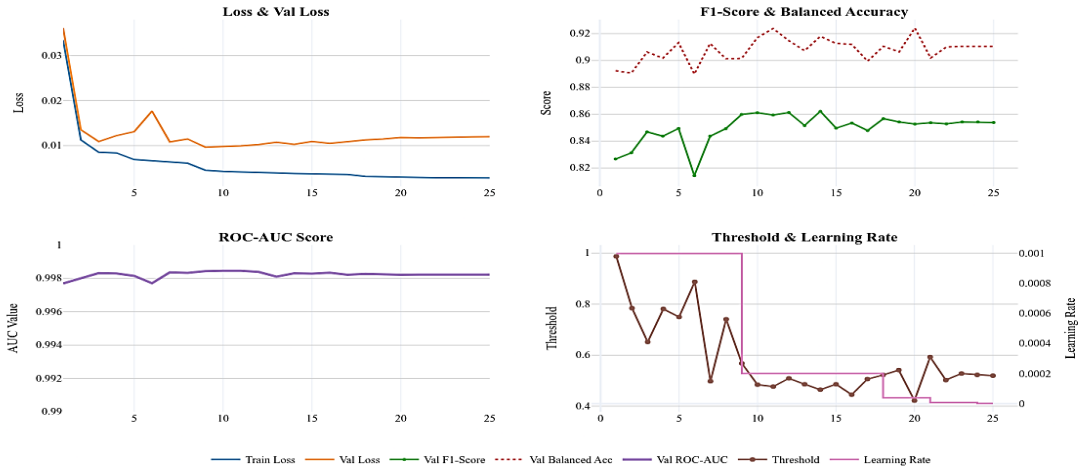
TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

3. *Distribution Of Merchant Risk*: Fitur ini merepresentasikan pola distribusi *multimodal* (banyak puncak). Secara visual, model berhasil mengidentifikasi lokasi klaster risiko utama, yang ditandai dengan puncak densitas pada rentang nilai 0,4; 0,5; dan 0,8. Meskipun kurva sintetis cenderung memiliki amplitudo yang lebih tajam (*over-sharpening*) dibandingkan data asli, struktur utama distribusi tetap terjaga. Hal ini menjamin bahwa informasi krusial mengenai kategori *merchant* berisiko tinggi tidak hilang selama proses sintesis.

Secara keseluruhan, analisis visual ini membuktikan bahwa data sintetis yang dihasilkan oleh CTGAN tidak sekadar melakukan memorisasi (menyalin data), melainkan berhasil mempelajari struktur probabilitas (*probability structure*) dari data latih secara mendalam.

3.2 *Model Training Analysis*

Kinerja pelatihan model *TabTransformer* dianalisis selama 25 *epoch* untuk memantau stabilitas metrik dan memastikan model mampu melakukan generalisasi dengan baik pada data validasi. Visualisasi lengkap mengenai dinamika pelatihan ini dapat dilihat pada Gambar 7.



Gambar 7.*Training History and Metrics Visualization*

Dinamika pembelajaran model dapat dijabarkan dalam 3 poin analisis berikut:

1. *Loss Convergence and Generalization Capability*: Pada grafik *Loss & Val Loss*, kurva *Train Loss* (garis biru) menunjukkan penurunan yang stabil, menandakan model berhasil mempelajari fitur data dengan baik. Sementara itu, *Validation Loss* (garis oranye) yang sempat fluktuatif di awal, mulai stabil setelah *epoch* ke-9. Jarak yang sempit antara kedua kurva di pertengahan proses menunjukkan kemampuan generalisasi yang optimal. Namun, perlu dicatat adanya sedikit kenaikan pada *Validation Loss* setelah *epoch* 20 (dari 0.0117 menjadi 0.0120), yang menjadi indikasi awal terjadinya *overfitting*.
2. *Clasification Performance Stability*: Dari sisi metrik evaluasi, model terbukti mampu menangani ketidakseimbangan data secara efektif. Hal ini terlihat dari *Balanced Accuracy* yang konsisten di atas 0.90 dan *ROC-AUC* yang stabil di kisaran 0.99. Selain itu, *F1-Score* mengalami peningkatan signifikan setelah *epoch* 7 dan mencapai performa puncaknya di pertengahan pelatihan, mengonfirmasi bahwa model dapat membedakan kelas *fraud* dan *non-fraud* dengan akurat.
3. *Optimization Strategy (Learning Rate and Threshold)*: Grafik *Threshold & Learning Rate* memperlihatkan efektivitas strategi *Step Decay*. Penurunan *Learning Rate* pada *epoch* 9 dan 18 terbukti berhasil menstabilkan pergerakan *Optimal Threshold*. Pada tahap akhir, nilai *threshold* bergerak stabil di rentang 0.44–0.54, menunjukkan bahwa model secara otomatis menyesuaikan sensitivitasnya untuk mendapatkan *F1-Score* terbaik.

Mengacu pada dinamika pelatihan tersebut, Epoch 14 dipilih sebagai titik *checkpoint* terbaik. Pada iterasi ini, model mencapai performa paling seimbang dengan *Validation Loss* yang rendah (0.0103) dan *F1-Score* tertinggi (0.8622). Pemilihan titik ini dilakukan untuk memaksimalkan akurasi sekaligus menghindari risiko *overfitting* yang mulai muncul pada *epoch-epoch* terakhir.

3.3 *Comparative Model Evaluation*

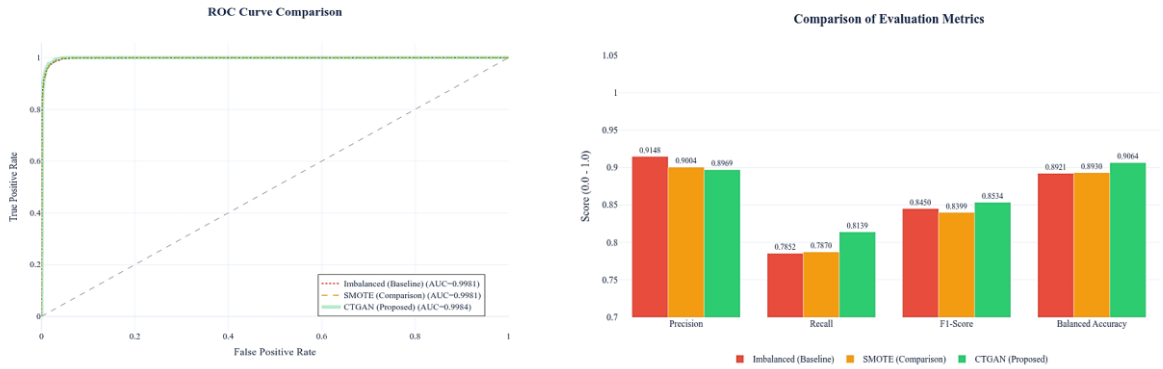
Evaluasi ini dilakukan untuk mengukur efektivitas penggunaan data sintetis (CTGAN) dalam meningkatkan performa model *TabTransformer*. Sebagai pembanding, model juga diuji terhadap dua skenario kontrol: data asli (*Imbalanced*) dan data yang diperbaiki dengan teknik konvensional (SMOTE). Ringkasan perbandingan kinerja ketiga model disajikan pada Tabel 7.

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

Tabel 7. Komparasi Performa Model

Data	Optimal Threshold	Precision	Recall	F1-Score	B-Accuracy	ROC-AUC
Imbalanced	0.6555	0.9148	0.7852	0.8450	0.8921	0.9981
SMOTE	0.9829	0.9004	0.7870	0.8399	0.8930	0.9981
CTGAN	0.4645	0.8969	0.8139	0.8534	0.9064	0.9984

Untuk melihat perbandingan yang lebih jelas, hasil evaluasi divisualisasikan menggunakan kurva ROC pada Gambar 8(a) dan diagram batang pada 8(b).



(a) (b)

Gambar 8. Hasil evaluasi: (a) *ROC Curve Comparation* (b) *Evaluation Metrics Comparation*

Temuan eksperimental tersebut mengindikasikan pola kinerja yang dapat dijabarkan sebagai berikut:

1. *Limitations of Conventional Oversampling* (SMOTE): Pada skenario SMOTE, terlihat adanya anomali kinerja. Meskipun teknik ini bertujuan menyeimbangkan data, *F1-Score* yang dihasilkan (0.8399) justru sedikit lebih rendah dibandingkan data asli/imbalanced (0.8450). Selain itu, kalibrasi model SMOTE terlihat kurang baik, ditandai dengan *Optimal Threshold* yang sangat tinggi (0.9829). Angka ekstrem ini mengindikasikan bahwa model cenderung "ragu-ragu" akibat adanya *noise* atau tumpang tindih (*overlapping*) data di area batas keputusan. Interpolasi linear yang dilakukan SMOTE cenderung menciptakan sampel sintetis di ruang kosong yang tidak realistis secara semantik, sehingga mengaburkan *decision boundary* yang sebenarnya. Akibatnya, model kehilangan sensitivitasnya dan membutuhkan probabilitas nyaris 100% untuk berani memprediksi *fraud*, kondisi yang sangat berisiko dalam sistem keamanan finansial.
2. *Superiority of Generative Approach* (CTGAN): Sebaliknya, pendekatan menggunakan CTGAN menunjukkan hasil yang paling optimal. Model ini unggul pada metrik *Recall* (0.8139) dan *F1-Score* (0.8534). Tingginya nilai *Recall* ini sangat penting dalam deteksi *fraud* untuk meminimalkan lolosnya transaksi curang (*False Negative*). Berbeda dengan SMOTE, model CTGAN memiliki *threshold* yang stabil di angka 0.4645 (mendekati 0.5), yang membuktikan bahwa data sintetis mampu memperjelas batasan antar kelas tanpa merusak distribusi data aslinya. Keunggulan ini berasal dari kemampuan CTGAN dalam mempelajari *manifold* data yang kompleks melalui adversarial training, sehingga sampel yang dihasilkan tetap berada dalam distribusi fitur yang valid namun menambah variasi yang diperlukan oleh model untuk belajar.
3. *Operational and Financial Implications*: Dari perspektif operasional perbankan, penggunaan model berbasis CTGAN menawarkan efisiensi yang signifikan. Dengan nilai Recall yang tinggi, bank dapat mengurangi risiko kerugian finansial akibat chargeback fraud yang tidak terdeteksi. Meskipun threshold yang lebih rendah mungkin sedikit meningkatkan jumlah False Positive (transaksi sah yang dicurigai), biaya operasional untuk verifikasi manual melalui SMS atau telepon jauh lebih rendah dibandingkan kerugian reputasi dan finansial akibat fraud yang lolos. Selain itu, stabilitas model ini menjamin bahwa sistem tidak perlu dikalibrasi ulang terlalu sering, sehingga mengurangi downtime dan biaya maintenance sistem deteksi fraud dalam jangka panjang.

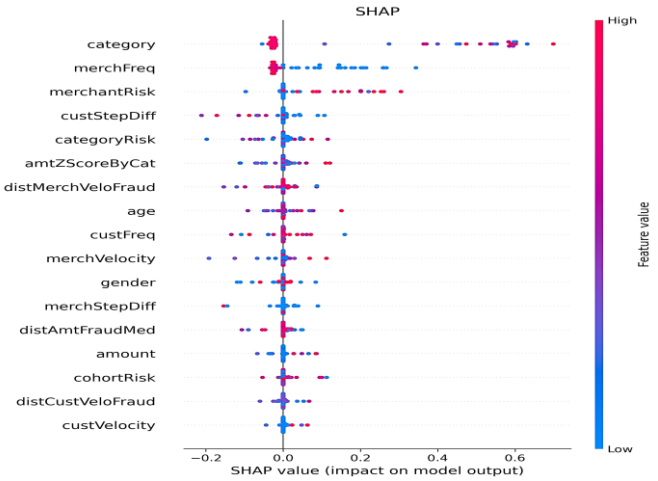
Temuan ini mengisi celah fundamental dalam literatur dengan mengalihkan fokus dari pendekatan Model-Centric ke Data-Centric AI. Studi terdahulu seperti Dharmana dkk. (2024) melalui teknik ADASYN hanya mencatatkan *F1-Score* sebesar 81%, sementara arsitektur kompleks *Hybrid Autoencoder-Transformer* milik Priatna dkk. (2025) stagnan pada *F1-Score* 80% dan *Recall* 74%. Hal ini mengindikasikan bahwa algoritma *Deep Learning* canggih sekalipun tidak akan optimal tanpa pengayaan data yang kontekstual.

Berbeda secara signifikan, penelitian ini membuktikan bahwa kualitas fitur lebih krusial daripada kompleksitas arsitektur. Melalui integrasi *Feature Engineering* berbasis domain (*velocity* dan *risk scoring*) serta dukungan data sintetis CTGAN, model yang diusulkan berhasil mencapai *F1-Score* 85,34% dan *Recall* 81,39%. Peningkatan performa dibandingkan *baseline* Priatna dkk. (2025) ini mengonfirmasi secara empiris bahwa strategi pengayaan data adalah kunci utama dalam melampaui batasan kinerja deteksi anomali finansial.

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

3.4 *Model Interpretability Analysis* (SHAP)

Analisis lanjutan menggunakan metode SHAP (*SHapley Additive exPlanations*) dilakukan untuk menginterpretasikan keputusan model *TabTransformer* yang bersifat *black-box*. Analisis ini bertujuan untuk memvalidasi apakah fitur-fitur hasil rekayasa (*feature engineering*) benar-benar memberikan kontribusi signifikan terhadap deteksi *fraud*. Visualisasi global mengenai kontribusi fitur disajikan pada Gambar 10.



Gambar 9.SHAP *Summary Plot*

Interpretasi perilaku model melalui *summary plot* tersebut dapat dijabarkan dalam 3 poin utama berikut:

1. *Dominance of Engineered Features*: Visualisasi tersebut mengonfirmasi efektivitas strategi rekayasa fitur dalam penelitian ini. Terlihat bahwa daftar 10 fitur dengan pengaruh terbesar didominasi oleh fitur turunan (*engineered features*), seperti *merchantRisk*, *custStepDiff*, dan *categoryRisk*. Dominasi ini menjadi bukti empiris bahwa variabel risiko yang diekstraksi melalui proses feature engineering memberikan informasi yang jauh lebih berharga bagi model dibandingkan sekadar menggunakan data transaksional mentah.
2. *Behavioral Patterns and Risk Correlation*: Fitur *categoryRisk* dan *merchantRisk* memperlihatkan korelasi linear positif terhadap prediksi. Sebaran titik merah (nilai risiko tinggi) yang terkonsentrasi di sisi kanan (nilai SHAP positif) menegaskan bahwa semakin tinggi skor risiko historis, semakin besar probabilitas transaksi dideteksi sebagai fraud. Sebaliknya, pada fitur *merchFreq* ditemukan pola hubungan terbalik. Titik biru (frekuensi rendah) yang berada di area positif mengindikasikan bahwa model cenderung mencurigai transaksi yang dilakukan pada merchant yang jarang atau baru pertama kali dikunjungi oleh nasabah.
3. *Statistical Anomalies as Fraud Signals*: Kontribusi signifikan juga ditunjukkan oleh fitur statistik *amtZScoreByCat*. Pola sebaran data menunjukkan bahwa model sangat sensitif terhadap deviasi nilai transaksi. Transaksi dengan nominal yang menyimpang jauh dari rata-rata kategori (*Z-Score* tinggi) secara konsisten mendorong prediksi ke arah *fraud*. Hal ini membuktikan bahwa model berhasil menangkap logika deteksi anomali berbasis statistik secara akurat.

Secara keseluruhan, analisis SHAP ini tidak hanya memvalidasi performa teknis model, tetapi juga menjamin aspek akuntabilitas sistem. Dalam industri finansial yang terikat regulasi ketat, kemampuan untuk menjelaskan alasan di balik pemblokiran transaksi (*Explainable AI*) adalah syarat mutlak. Fakta bahwa model *TabTransformer* mendasarkan keputusannya pada indikator logis—seperti anomali nilai transaksi dan profil risiko—menegaskan bahwa performa tingginya bukan disebabkan oleh *noise*, melainkan pemahaman mendalam terhadap pola kejahatan finansial. Dominasi fitur rekayasa dalam plot ini sekaligus menjawab hipotesis penelitian bahwa kualitas *Feature Engineering* memegang peran sentral dalam keberhasilan klasifikasi kejahatan finansial.

Sinergi antara CTGAN, *TabTransformer*, dan SHAP dalam penelitian ini menawarkan kerangka kerja yang komprehensif untuk ekosistem keamanan perbankan digital. CTGAN menyelesaikan masalah di sisi hulu (ketersediaan data), *TabTransformer* memberikan performa tinggi di sisi pemrosesan (akurasi deteksi), dan SHAP memberikan validitas di sisi hilir (transparansi keputusan). Temuan ini mengimplikasikan bahwa masa depan deteksi fraud tidak bisa lagi hanya bergantung pada satu algoritma klasifikasi semata, melainkan membutuhkan orkestrasi *pipeline* cerdas yang menangani siklus hidup data dari sintesis hingga interpretasi.

4. KESIMPULAN

Penelitian ini secara komprehensif menyimpulkan bahwa pergeseran paradigma dari pendekatan *Model-Centric* menuju Data-Centric AI merupakan kunci fundamental untuk mengatasi tantangan klasifikasi *fraud* pada data dengan ketimpangan ekstrem. Integrasi strategis antara rekayasa fitur berbasis domain (*feature engineering*), sintesis data generatif (CTGAN), dan arsitektur TabTransformer terbukti mampu membentuk model deteksi yang jauh lebih cerdas

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

dibandingkan metode konvensional. Berbeda dengan studi terdahulu yang cenderung stagnan akibat ketergantungan pada interpolasi linear sederhana (SMOTE/ADASYN), pendekatan hibrida ini berhasil mereplikasi distribusi probabilitas gabungan yang kompleks. Hal ini dibuktikan secara empiris melalui capaian kinerja superior dengan *F1-Score* sebesar 85,34% dan *Recall* 81,39%, yang mengindikasikan bahwa model mampu mengenali pola anomali secara presisi tanpa mengorbankan sensitivitas. Secara implikasi praktis, temuan ini memberikan kontribusi strategis bagi ekosistem perbankan digital. Kemampuan model dalam menekan angka *False Negative* berdampak langsung pada minimalisasi kerugian finansial akibat transaksi curang yang tidak terdeteksi. Lebih jauh lagi, integrasi metode interpretasi SHAP menyediakan transparansi keputusan yang akuntabel, sehingga mempermudah proses audit kepatuhan dan meningkatkan efisiensi operasional tim analis dalam memverifikasi peringatan dini. Namun, penelitian ini memiliki keterbatasan utama pada penggunaan dataset simulasi yang bersifat statis, sehingga belum sepenuhnya menangkap fenomena perubahan perilaku serangan yang dinamis (*concept drift*) yang kerap terjadi di dunia nyata. Oleh karena itu, penelitian selanjutnya sangat direkomendasikan untuk memvalidasi kerangka kerja ini pada data transaksi riil berskala besar serta menerapkan mekanisme pembelajaran berkelanjutan (*continuous learning*) agar sistem pertahanan tetap adaptif dan relevan dalam menghadapi evolusi modus kejahatan finansial yang terus berkembang.

REFERENCES

Alshawi, B. (2023). Utilizing GANs for credit card fraud detection: A comparison of supervised learning algorithms. *Engineering, Technology & Applied Science Research*, *13*(6), 12264–12270. https://doi.org/10.48084/etasr.6434

Assabil, J. J. (2024). Credit card fraud detection using machine learning algorithms: A comparative study of six models. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(23s), 862–875. https://ijisae.org/index.php/IJISAE/article/view/7040

Bank Indonesia. (2024, April 27). *Laporan kebijakan moneter triwulan IV 2024*. Departemen Kebijakan Ekonomi dan Moneter. https://www.bi.go.id/id/publikasi/laporan/Pages/Laporan-Kebijakan-Moneter-Triwulan-I-2024.aspx

Bonde, L., & Bichanga, A. K. (2025). Improving credit card fraud detection with ensemble deep learning-based models: A hybrid approach using SMOTE-ENN. *Journal of Computing Theories and Applications*, *2*(3), 383–394. https://doi.org/10.62411/jcta.12021

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

Conteh, A. A., & Zhou, J. (2025). Credit card fraud detection with imbalanced small data using TabTransformer and cost-sensitive learning. Dalam G. Xu, W. Zhou, J. Zhang, Y. Zhang, & Y. Jia (Ed.), *Cyberspace Simulation and Evaluation* (Vol. 2422, hlm. 35–50). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-4509-1_3

Dharmana, I. W., Gunadi, I. G. A., & Dewi, L. J. E. (2024). Deteksi transaksi fraud kartu kredit menggunakan oversampling ADASYN dan seleksi fitur SVM-RFECV. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *11*(1), 125–134. https://doi.org/10.25126/jtiik.20241117640

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. http://www.deeplearningbook.org

Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). *TabTransformer: Tabular data modeling using contextual embeddings* (No. arXiv:2012.06678). arXiv. https://doi.org/10.48550/arXiv.2012.06678

Ibrahim, M. M., & Alfauzan, S. (2025). Analisis kinerja model machine learning untuk mendeteksi transaksi fraud pada sistem pembayaran online. *Jurnal Ilmiah Nusantara*, *2*(3), 35–49. https://doi.org/10.61722/jinu.v2i3.4276

Lopez-Rojas, E. A., & Axelsson, S. (2014, September 10). *BankSim: A bank payment simulation for fraud detection research*. 26th European Modeling and Simulation Symposium, EMSS 2014. https://www.researchgate.net/publication/265736405_BankSim_A_Bank_Payment_Simulation_for_Fraud_Detection_Research

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*. https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Nugraha, R. A., Pardede, H. F., & Subekti, A. (2022). Oversampling based on generative adversarial networks to overcome imbalance data in predicting fraud insurance claim. *Kuwait Journal of Science*. https://doi.org/10.48129/kjs.splml.19119

Odeyemi, O., Mhlongo, N. Z., Nwankwo, E. E., & Oluwatobi Timothy Soyombo. (2024). Reviewing the role of AI in fraud detection and prevention in financial services. *International Journal of Science and Research Archive*, *11*(1), 2101–2110. https://doi.org/10.30574/ijsra.2024.11.1.0279

Padhi, I., Schiff, Y., Melnyk, I., Rigotti, M., Mroueh, Y., Dognin, P., Ross, J., Nair, R., & Altman, E. (2021). Tabular transformers for modeling multivariate time series. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3565–3569. https://doi.org/10.1109/ICASSP39728.2021.9414142

Patil, T. (2021). *Credit card fraud detection using Conditional Tabular Generative Adversarial Networks (CT-GAN) and supervised machine learning techniques* [Master's thesis, National College of Ireland]. https://norma.ncirl.ie/5209/1/tusharvasantpatil.pdf

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1308-1321
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9056

Priatna, W., Prasetyo, S. Y. J., Wijono, S., Maria, E., & Manongga, D. (2025). Deteksi Anomali dalam Penipuan E-commerce Menggunakan Hybrid Autoencoder-Transformer Frameworks. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, *11*(1), 33–40. https://doi.org/10.26418/jp.v11i1.82330

Putri, K. A. (2025, Februari 11). OJK terima 42.257 laporan penipuan, total kerugian korban tembus Rp700,2 M. *Infobanknews*. https://infobanknews.com/ojk-terima-42-257-laporan-penipuan-total-kerugian-korban-tembus-rp7002-m/

World Bank Group. (2023, Oktober). *Fraud risks in fast payments*. https://fastpayments.worldbank.org/sites/default/files/2023-10/Fraud%20in%20Fast%20Payments_Final.pdf

Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using Conditional GAN. *Advances in Neural Information Processing Systems*, *32*. https://proceedings.neurips.cc/paper_files/paper/2019/hash/254ed7d2de3b23ab10936522dd547b78-Abstract.html

Yuhertiana, I., & Amin, A. H. (2024). Artificial intelligence driven approaches for financial fraud detection: A systematic literature review. *KnE Social Sciences*. https://doi.org/10.18502/kss.v9i20.16551