Journal of Applied Informatics and Computing (JAIC)
Vol.8, No.2, December 2024, pp. 580~588
e-ISSN: 2548-6861

Optimized Machine Learning Model for Credit Card Fraud Detection Using Smote-Tomek and Feature Engineering

Mochammad Abdurrochman Ari Wibowo [1]*, De Rosal Ignatius Moses Setiadi [2]**
* Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, 50131, Indonesia
** Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, 50131, Indonesia
arie35411@gmail.com [1], moses@dsn.dinus.ac.id [2]

Article Info

Article history:

Received 2024-10-23
Revised 2024-11-01
Accepted 2024-11-04

Keyword:

Credit card,
Feature Engineering,
Machine Learning,
Oversampling,
SMOTE-Tomek.

ABSTRACT

In today's digital economy, credit cards are indispensable, with both usage and fraud incidents increasing significantly in recent years. According to data from the Payment System and Financial Market Infrastructure Statistics (SPIP), the volume of credit card transactions in Indonesia reached 28,360 in May 2022, higher than the 23,452 transactions recorded in May 2021. Credit card fraud can be detected using machine learning models that analyze suspicious transaction histories. However, challenges arise due to the imbalance in credit transaction data, where models tend to favor the majority class, in this case, non-fraudulent transactions, while the minority class (fraudulent transactions) is more critical to identify. This study proposes data balancing techniques using SMOTE-Tomek, polynomial feature engineering, and feature selection with the SelectKBest function to optimize the performance of the random forest model. The dataset, publicly available on Kaggle, consists of 283,807 transactions, including 492 labeled as fraudulent and 284,315 as non-fraudulent. The training and validation process employed 5-fold cross-validation to ensure reliable results. Findings reveal that the random forest model outperformed three other models—gradient boosting, decision tree, and XGBoost—with recall and F1-score exceeding 99%. It can be concluded that the integration of feature engineering, feature selection, and SMOTE-Tomek significantly enhances the performance of machine learning models in detecting credit card transaction fraud.

I. PENDAHULUAN

Pada era sekarang, kemajuan teknologi sudah sangat berkembang dengan pesat dan mempunyai dampak yang signifikan terutama berdampak pada sektor keuangan, sehingga menyebabkan adopsi yang luas platform perdagangan elektronik. Dengan adanya perkembangan ini menunjukan bahwa perlunya dunia yang lebih digital dan semakin memperluas industri e-commerce [1]. Salah satu permasalahan utama yang terkait dengan e-commerce modern adalah tingginya kasus penipuan kartu kredit [2]. Penipuan kartu kredit adalah masalah serius yang mengakibatkan kerugian financial secara signifikan bagi konsumen atau lembaga perusahaan keuangan. Dengan peningkatan penggunaan kartu kredit sebagai alat pembayaran di era modern ini, resiko penipuan juga meningkat[3]. Meningkatnya tingkat penipuan kartu kredit dikaitkan dengan perluasan e-commerce dan peningkatan transaksi online. Berdasarkan data Statistik Sistem Pembayaran dan Infrastruktur Pasar Keuangan (SPIP) yang diterbitkan oleh Bank Indonesia pada juni 2022 dapat diketahui bahwa jumlah kartu kredit yang beredar di indonesia mengalami peningkatan sejak 10 tahun terakhir, dimana per Mei 2022 jumlah kartu kredit yang beredar tercatat sebanyak 16.588.263 unit, bertambah hampir dua juta unit dibandingkan dengan jumlah pada tahun 2012 yang sebanyak 14.817.168 unit. Sedangkan untuk volume transaksi menggunakan kartu kredit di Indonesia pada bulan Mei 2022 mencapai 28.360 transaksi, lebih tinggi dibandingkan volume pada Mei 2021 sejumlah 23.452 transaksi[4].

Penipuan kartu kredit dapat dibagi menjadi 2 jenis: penipuan kartu internal dan penipuan kartu eksternal. Penipuan kartu internal bertujuan untuk menggelapkan uang tunai. Biasanya, ini adalah kolusi antara pedagang dan pemegang kartu, menggunakan transaksi palsu untuk menggelapkan uang tunai bank. Penipuan kartu eksternal terutama diwujudkan dengan menggunakan kartu kredit yang dicuri, palsu, atau tiruan untuk berbelanja, atau menggunakan kartu untuk mendapatkan uang tunai dalam bentuk yang disamarkan, seperti membeli barang mahal dengan volume kecil atau komoditas yang dapat dengan mudah diubah menjadi uang tunai [5].

Untuk mengatasi masalah ini, diperlukan alogaritma klasifikasi yang efektif yang mampu mengidentifikasi transaksi yang mencurigakan dengan akurasi tinggi dan tingkat false positive yang rendah. Beberapa sistem berbasis pembelajaran mesin telah dikembangkan untuk mendeteksi penipuan kartu kredit. Misalnya, Malik et al. [6] mempelajari pengembangan model hybird dalam credit card fraud detection. Model hybird dicapai dengan menggabungkan berbagai alogaritma pembelajaran mesin, termasuk gradient boosting, random forest, decision tree dan XGBoost. Namun, ketidakseimbangan data antar kelas merupakan tantangan yang sering dihadapi dalam menggunakan data credit card.

Berbagai teknik dapat digunakan untuk mengatasi masalah ini, termasuk pengembangan alogaritma yang mempertimbangkan nilai kelas positif, atau dengan memodifikasi dataset untuk meningkatkan perfoma klasifikasi[7]. Ileberi et al.[8] dan Mujahid et al.[9] penggunaan teknik oversampling untuk mengatasi ketidakseimbangan data penipuan kartu kredit. Selain itu Mujahid et al.[9] juga menggunakan teknik rekayasa fitur untuk meningkatkan kinerja model pembelajaran mesin [8]. Kombinasi teknik oversampling dan rekayasa fitur juga berhasil meningkatkan performa pembelajaran mesin pada bidang medis [10]. Pada bidang deteksi intrusi pada penelitian [11] hanya digunakan teknik rekayasa fitur. Metode seleksi fitur juga telah banyak diterapkan dalam meningkatkan performa pemebelajaran mesin, seperti pada penelitian [12][13][14][15]. Sedangkan pada penelitian [16][17] dikombinasikan teknik balancing dan seleksi fitur.

Namun, ketidakseimbangan data menjadi tantangan signifikan dalam mendeteksi penipuan kartu kredit karena data transaksi biasanya didominasi oleh transaksi non-penipuan, sehingga model prediksi cenderung bias terhadap kelas mayoritas. Akibatnya, model mengabaikan kelas minoritas, yaitu transaksi penipuan, yang mana menjadi fokus utama dalam penelitian ini. Untuk mengatasi masalah tersebut, penelitian ini mengkombinasikan tiga teknik yaitu SMOTE-Tomek untuk mengatasi ketidakseimbangan data, transformasi polinomial untuk menambah variasi fitur, dan seleksi fitur untuk memilih atribut fitur yang terbaik dalam mendeteksi penipuan. Kombinasi tiga metode ini belum banyak digunakan, sehingga diharapkan dapat meningkatkan performa model terutama dalam hal akurasi, presisi, dan recall pada kelas penipuan. Dengan demikian, pendekatan ini bertujuan meningkatkan keandalan model dalam mendeteksi transaksi penipuan kartu kredit.

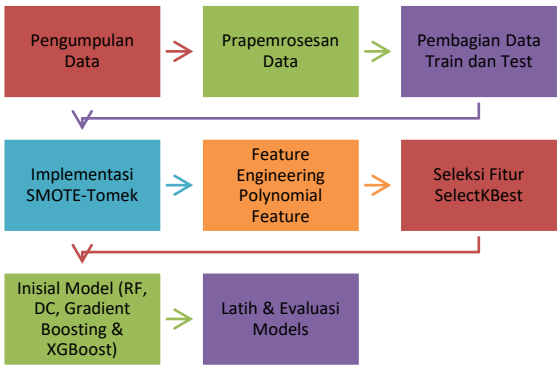
Synthetic Minority Over-sampling Technique berfungsi menciptakan sampel sintetis dari kelas minoritas untuk menyeimbangkan distribusi kelas. Sementara itu, Tomek links mengidentifikasi dan menghapus pasangan sampel dari kelas yang berbeda yang saling berdekatan. Kombinasi SMOTE dan Tomek links menciptakan proses oversampling dan undersampling untuk membentuk dataset yang seimbang tanpa tumpang tindih antara kelas. Polynomial Features berfungsi untuk membuat interaksi antar fitur dan meningkatkan keragaman fitur dengan menambahkan kolom-kolom yang merupakan hasil perkalian dari fitur awal dan seleksi fitur (*SelectKBest*) berfungsi untuk mengurangi kompleksitas model dengan hanya menggunakan fitur yang paling relevan terhadap variabel target.

Berdasarkan literatur yang telah dibahas, tujuan dari penelitian ini adalah untuk menganalisis berbagai metode machine learning yang digunakan dalam mengklasifikasikan dataset kartu kredit. Penelitian ini juga bertujuan untuk menganalisis pengaruh teknik prapemrosesan, seperti seleksi fitur dan feature engineering, terhadap kinerja metode klasifikasi dalam menghadapi dataset kartu kredit yang tidak seimbang. Selain itu, penelitian ini berfokus pada pengembangan solusi untuk mengatasi masalah ketidakseimbangan kelas dalam dataset kartu kredit dengan menggunakan teknik SMOTE-Tomek.

II. METODE PENELITIAN

A. *Pengumpulan Data*

Penelitian ini menggunakan kumpulan data kartu kredit yang berisi transaksi yang dilakukan oleh pemegang kartu Eropa pada bulan September 2013 yang tersedia di kaggle [18].



Gambar 1. Metode Yang Diusulkan

Dataset berisi 283.807 transaksi di antaranya 492 transaksi diberi label sebagai penipuan dan 284,315 diberi label sebagai tidak penipuan. Datasetnya sangat tidak seimbang, dengan hanya 0,172% yang diberi label sebagai transaksi penipuan. Sebagian besar fitur diubah menjadi variabel numerik

menggunakan analisis komponen utama (PCA) karena masalah kerahasiaan, dan nama fitur dianonimkan sebagai V1, V2, V3, . . . , dan V28, tidak termasuk fitur "Waktu" dan "Jumlah". Fitur "Kelas" adalah variabel target, dan memiliki nilai 1 dan 0, masing-masing mewakili transaksi penipuan dan non-penipuan.

V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
0.517232	-0.035049	-0.465211	0.320198	0.044519	0.177840	0.261145	-0.143276	0.00	1
0.661696	0.435477	1.375966	-0.293803	0.279798	-0.145362	-0.252773	0.035764	529.00	1
-0.294166	-0.932391	0.172726	-0.087330	-0.156114	-0.542628	0.039566	-0.153029	239.93	1
0.573574	0.176968	-0.436207	-0.053502	0.252405	-0.657488	-0.827136	0.849573	59.00	1
-0.379068	-0.704181	-0.656805	-1.632653	1.488901	0.566797	-0.010016	0.146793	1.00	1
...	...	...	...	...	...	...	...	...	...
-0.426232	-0.839282	0.453846	-0.834638	-0.791865	-0.600372	0.077427	-0.028646	110.87	0
0.419152	-0.858150	0.536884	-0.356567	-0.203882	0.158595	0.097690	0.104090	4.49	0
0.016397	0.212432	-0.260651	-0.047875	-0.045265	0.594097	0.242007	0.201287	19.99	0
0.319091	1.049602	-0.272283	-0.412798	0.798130	0.134963	-0.086414	-0.100336	14.47	0
-0.089678	-0.802460	0.212016	-0.985291	-0.869253	-0.361544	-0.016930	0.014855	384.36	0

Gambar 2. Sampel Dataset Credit Card

B. Prapemrosesan Data

Pada tahap ini, akan dilakukan proses persiapan data sebelum akhirnya dataset siap digunakan untuk diterapkan pada model. Tahapan pertama adalah cek apakah ada nilai yang hilang dalam data. Kedua, cek apakah ada duplikasi data. Selanjutya, melakukan normalisasi dan standarisasi data.

C. Pemisahan Data

Proses klasifikasi merupakan pembelajaran tanpa pengawasan sehigga perlu proses pelatihan dan validasi. Langkah ini memastikan bahwa model dapat di evaluasi secara akurat dan tidak overfitting. Pada penelitian ini kami menggunakan menggunakan komposisi 80% untuk pelatihan data dan 20% untuk pengujian data.

D. Menerapakan Teknik Smote-Tomek

SMOTE (Synthetic Minority Over-Sampling Technique) dan Tomek links adalah dua teknik resampling yang sering digunakan untuk mengatasi masalah ketidakseimbangan kelas dalam dataset. SMOTE bertujuan untuk meningkatkan jumlah sampel dalam kelas minoritas, sementara Tomek links bertujuan untuk menghapus pasangan sampel yang berdekatan dari kelas yang berbeda untuk membersihkan data dari tumpang tindih antara kelas [12]. Adapun untuk formula rumus dari Smote sebagai berikut:

$$x_{new} = x_i + (x_j - x_i) \times \lambda$$

Dimana x_i x_j adalah sampel minoritas terdekat dan λ adalah nilai acak antara 0 dan 1. Setelah menerapkan SMOTE, teknik Tomek Links digunakan untuk memperjelas batas keputusan antar kelas dengan menghapus pasangan sampel yang berdekatan tetapi berasal dari kelas yang berbeda, sehingga mengurangi tumpang tindih dan meningkatkan kejelasan batas kelas. Kombinasi SMOTE dan Tomek Links, yang dikenal sebagai SMOTE-Tomek, diterapkan hanya pada data pelatihan untuk memastikan dataset yang lebih seimbang.

E. Menerapkan Teknik Feature Engineering

Rekayasa fitur adalah tugas utama dalam persiapan data untuk pembelajaran mesin. Rekayasa fitur melibatkan penerapan fungsi transformasi seperti operator aritmatika dan agregat operator pada fitur yang diberikan untuk menghasilkan fitur baru. Transformasi membantu menskalakan fitur atau mengubah hubungan non-linear antara fitur dan kelas target menjadi hubungan linear, yang lebih mudah dipelajari [13]. Pada proses ini, data dipersiapkan dan ditransformasikan sehingga lebih mudah dipahami oleh algoritma untuk membantu model dalam membuat prediksi yang lebih akurat.

F. Inisialisasi Model

Pada tahapan ini, dataset yang telah melalui tahap pra-pemrosesan dan resampling akan melalui proses klasifikasi. Dalam penelitian ini, beberapa model supervised machine learning seperti Random Forest, Gradient Boosting, Decision tree dan XGBoost akan digunakan sebagai alogaritma machine learning. Dalam dekade terakhir ini banyak penelitian menunjukan Supervised Machine Learning Algorithms Classification cenderung memberikan akurasi klasifikasi yang lebih baik [19].

Random Forest adalah algoritma dengan banyak pohon keputusan yang bekerja bersama sama. Setiap pohon memberikan prediksi dan hasilnya divoting untuk menentukan kelas yang paling banyak didukung sebagai prediksi akhir model. Algoritma ini tahan terhadap noise, dapat memproses data dengan variasi tinggi tanpa memerlukan seleksi fitur, dan mampu mengidentifikasi variabel penting [20].

Extreme Gradient Boosting (XGBoost) adalah algoritma berbasis pohon yang baru dan populer untuk klasifikasi data. XGBoost telah terbukti menjadi metode yang sangat efektif untuk klasifikasi data. XGBoost adalah sistem tree boosting yang sangat skalabel, digunakan dalam machine learning untuk tugas klasifikasi dan regresi [21].

Decision Tree adalah model hierarki pendukung keputusan yang menggunakan model keputusan seperti pohon dan kemungkinan konsekuensinya, termasuk hasil kejadian yang tidak disengaja, biaya sumber daya, dan utilitas. Ini adalah salah satu cara untuk menampilkan algoritma yang hanya berisi pernyataan kontrol bersyarat [22].

Gradient Boosting adalah teknik pembelajaran mesin berdasarkan boosting dalam ruang fungsional, di mana targetnya adalah pseudo-residual daripada residual tradisional yang digunakan dalam boosting tradisional. Teknik ini menghasilkan model prediksi dalam bentuk ensembel model prediksi lemah, yaitu model yang membuat asumsi sangat sedikit tentang data, yang biasanya berupa pohon keputusan sederhana [23].

G. Evaluasi Model

Sebagai tolak ukur, evaluasi dari kinerja model dalam penelitian ini akan menggunakan beberapa metrik penilaian seperti akurasi, presisi, recall, specificity dan F1-Score dan. Metrik-metrik ini dihitung berdasarkan hasil *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Akurasi adalah proporsi prediksi benar yang dibuat oleh model dari persamaan 1 berikut:

$$akurasi = \frac{(TP + TN)}{(TP + TN + FN + FN)}$$

Presisi adalah proporsi prediksi positif yang benar dari semua prediksi yang dikategorikan sebagai positif oleh model dari hasil persamaan 2 berikut:

$$presisi = \frac{Tp}{(TP + FP)}$$

Recall yang juga dikenal sebagai sensitivity adalah proporsi prediksi positif yang benar dari semua kasus positif yang sebenarnya dari hasil persamaan 2 berikut:

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

F1-Score adalah rata-rata harmonis antara presisi dan recall dari hasil persamaan 4 berikut:

$$\mathcal{F}1 = 2 \times \frac{Presisi \times Recall}{Presisi \times Recall}$$

Specificity adalah proporsi prediksi negatif yang benar dari semua prediksi yang dikategorikan sebagai negatif oleh model dari hasil persamaan 5 berikut:

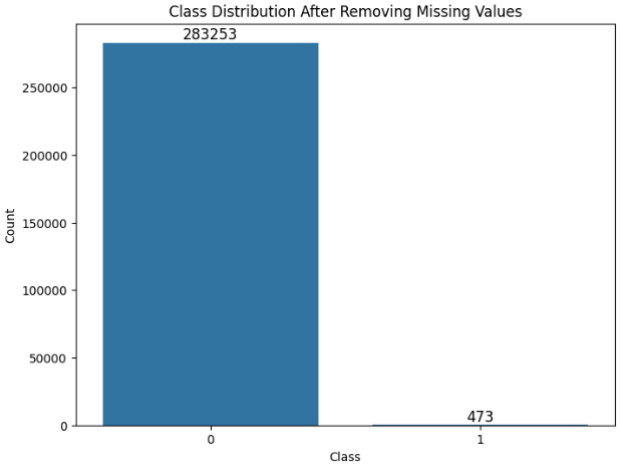
$$specifity = \frac{TN}{(TN + FP)}$$

III. HASIL DAN PEMBAHASAN

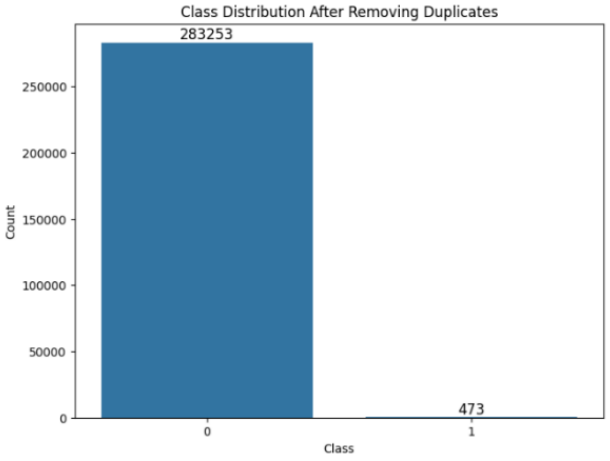
Dalam penelitian ini, Google Colab digunakan sebagai editor dan Python sebagai bahasa pemrograman, serta memanfaatkan GPU RTX 3050 untuk tugas komputasi. Selain itu, RAM 16 GB digunakan untuk memfasilitasi pemrosesan data dan pelatihan untuk model prediksi penipuan kartu kredit. Beberapa pustaka penting meliputi Sklearn dan imblearn.over_sampling. Bagian ini dibagi menjadi lima bagian utama: dua tahap praproses dataset, tahap rekayasa fitur, tahap validasi silang, dan pembahasan hasil klasifikasi.

A. Tahapan Prapemrosesan Pertama

Pada tahapan pertama untuk memastikan data konsisten di lakukan preprocessing data sebelum teknik oversampling diterapkan. Pada tahapan pertama ini dilakukan cek nilai yang hilang, hasil pengecekan dapat dilihat pada Gambar 3. Selanjutnya, melakukan pengecekan apakah ada data yang terduplikasi dalam dataset. Hasil pengecekan dapat dilihat pada Gambar 4. Serta melakukan normalisasi data yang hasilnya dapat dilihat pada gambar 6.



Gambar 1. Hasil Cek Nilai Hilang



Gambar 2. Hasil Hapus Duplikasi Data

Sample of data before normalization:

	Time	V1	V2	V3	V4	V5	V6	V7
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941

	V8	V9	...	V21	V22	V23	V24	V25
0	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.128539
1	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846	0.167170
2	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281	-0.327642
3	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175575	0.647376
4	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141267	-0.206010

	V26	V27	V28	Amount	Class
0	-0.189115	0.133558	-0.021053	149.62	0
1	0.125895	-0.008983	0.014724	2.69	0
2	-0.139097	-0.055353	-0.059752	378.66	0
3	-0.221929	0.062723	0.061458	123.50	0
4	0.502292	0.219422	0.215153	69.99	0

[5 rows x 31 columns]

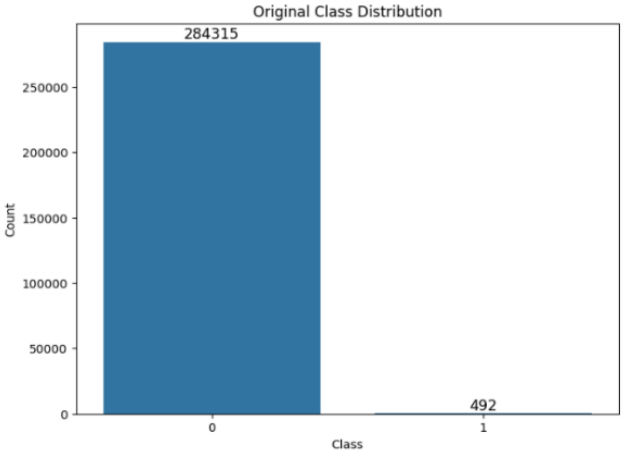
Gambar 3. Hasil Sebelum Normalisasi

```
Sample of data after normalization:
       Time        V1        V2        V3        V4        V5        V6  \
0 -1.996823 -0.701082 -0.041687  1.680101  0.976623 -0.247020  0.348012
1 -1.996823  0.608792  0.164138  0.109279  0.318998  0.042258 -0.060980
2 -1.996802 -0.700336 -0.811337  1.174270  0.270648 -0.366756  1.352655
3 -1.996802 -0.499064 -0.109972  1.187383 -0.608355 -0.008814  0.937245
4 -1.996781 -0.597606  0.535539  1.025470  0.287092 -0.297036  0.072873

         V7        V8        V9  ...       V21       V22       V23       V24  \
0  0.193700  0.084434  0.333534  ... -0.024777  0.383483 -0.177444  0.110157
1 -0.065656  0.072903 -0.231703  ... -0.311372 -0.881454  0.162081 -0.561503
2  0.643223  0.210788 -1.381169  ...  0.343094  1.065068  1.457772 -1.138484
3  0.192079  0.320843 -1.264664  ... -0.149093  0.007299 -0.305465 -1.941446
4  0.481517 -0.228725  0.747917  ... -0.012516  1.101780 -0.220709  0.232904

        V25       V26       V27       V28    Amount  Class
0  0.247059 -0.392622  0.333033 -0.065850  0.244200      0
1  0.321175  0.260854 -0.027154  0.043219 -0.342584      0
2 -0.628161 -0.288861 -0.144325 -0.183824  1.158900      0
3  1.242487 -0.460694  0.154039  0.185687  0.139886      0
4 -0.394800  1.041677  0.550001  0.654234 -0.073813      0

[5 rows x 31 columns]
```

Gambar 4. Hasil Setelah Normalisasi

B. Tahapan Prapemrosesan Kedua

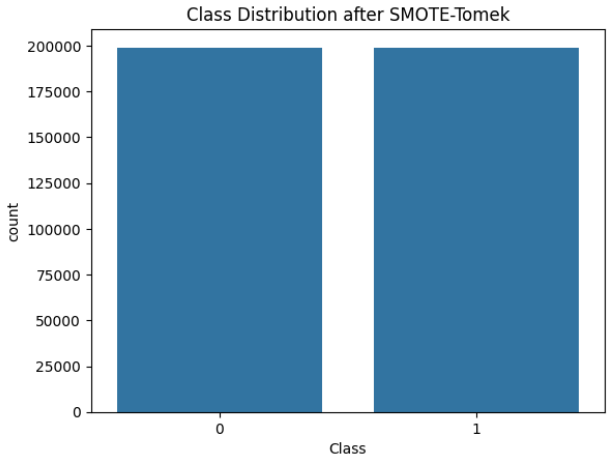
Terdapat proposisi data yang tidak seimbang dalam dataset credit card antara kelas (1) penipuan dan (0) tidak penipuan seperti yang ditampilkan pada Gambar 7.



Gambar 5. Distribusi Data Original

Dataset credit card memiliki ketidakseimbangan data yang cukup signifikan antara kelas 0 dan kelas 1. Terdapat 284.315 baris data pada kelas 0 (tidak penipuan), sedangkan pada kelas 1 (penipuan) hanya 492 baris data. Teknik oversampling digunakan untuk meningkatkan data kelas minoritas agar seimbang dengan data kelas mayoritas.

Pada teknik oversampling ini yaitu penerapan Smote-Tomek dengan mengindentifikasi kelas minoritas dalam kumpulan data. SMOTE bekerja dengan mencari k tetangga terdekat untuk setiap data di kelas minoritas. Oversampling berbasis cluster pengganti yang pada saat yang sama akan menangani ketidakseimbangan antar kelas dan dalam kelas tidak seimbang [24]. Teknik Tomek diperkenalkan oleh Tomke, Tautan Tomek bekerja dengan menghapus kelas negatif dan kelas positif yang memiliki kesamaan karakteristik [25]. Kubat dan Matwin mengembangkan teknik pengambilan sampel yang selektif untuk kelemahan pengetahuan yang tidak seimbang dan diperkenalkan prosedur pembersihan data yang menggunakan tautan Tomek konstruksi untuk menghapus kelas negatif yaitu masalah batas [26]. Bisa dilihat pada Gambar 8 data yang lebih seimbang antara kelas minoritas dan mayoritas setalah dilakukan Smote-Tomek.



Gambar 6. Hasil Setelah Dilakukan Smote-Tomek

SMOTE-Tomek dipilih dalam penelitian credit card fraud detection dengan dataset yang tidak seimbang karena kombinasi antara teknik over-sampling dan under-samplingnya yang efektif. SMOTE menghasilkan sampel sintetis dari kelas minoritas secara interpolatif, menambah variasi data tanpa hanya menggandakan, sehingga meminimalkan risiko overfitting yang sering terjadi pada metode lain seperti Random Over Sampling (ROS). Sementara itu, Tomek Links menghapus data mayoritas yang tumpang tindih dengan minoritas, mengurangi noise dan meningkatkan akurasi model. Dibandingkan dengan ADASYN, yang cenderung berfokus pada titik-titik sulit di kelas minoritas, SMOTE-Tomek lebih unggul karena menghasilkan data yang lebih bersih dan mengurangi risiko memasukkan noise berlebih. Kombinasi ini membuat SMOTE-Tomek ideal dalam menghasilkan dataset yang seimbang dan lebih stabil untuk model deteksi fraud kartu kredit.

C. Tahapan Feature Engineering

Tahapan feature engineering dalam penelitian ini dimulai dengan One-Hot Encoding untuk mengubah fitur kategorikal menjadi format numerik. Selanjutnya, fitur polinomial ditambahkan menggunakan PolynomialFeatures untuk menangkap pola non-linear dalam data. Kemudian, semua fitur distandardisasi dengan menggunakan StandardScaler untuk memastikan setiap fitur berada dalam skala yang seragam. Terakhir, dilakukan seleksi fitur menggunakan SelectKBest untuk memilih k fitur terbaik yang paling relevan dengan variabel target dengan menggunakan skema K-10, yang mana akan memilih 10 fitur terbaik dari data input x_train berdasarkan hubungan dengan variabel target y_train. Tahapan Proses ini dapat meningkatkan kualitas data dan kinerja model dengan fokus pada fitur yang paling signifikan.

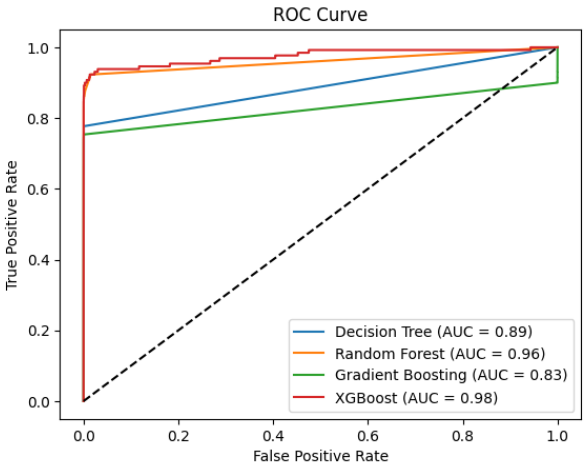
D. Tahapan Cross Validation

Tahapan cross validation ini bertujuan untuk mengevaluasi

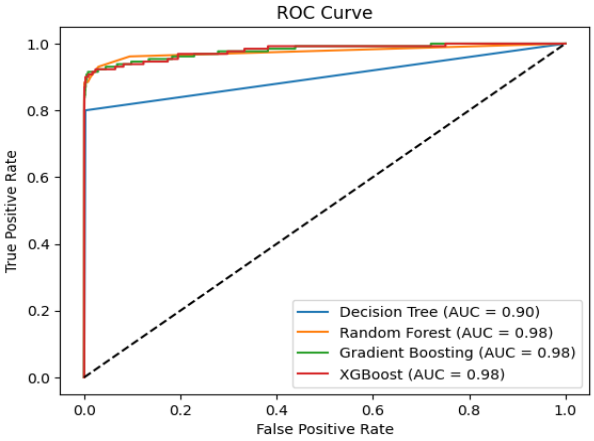
kinerja model secara lebih robust dengan meminimalkan kemungkinan overfitting. Metode yang diterapkan adalah *5-fold cross-validation* untuk membagi data pelatihan menjadi 5 lipatan, setiap iterasi model dilatih menggunakan empat lipatan sebagai data pelatihan dan satu lipatan digunakan sebagai data validasi. Prediksi kelas dan probabilitas prediksi dihasilkan menggunakan metode cross_val_predict dengan pendekatan "predict" dan "predict_proba". Skema 5-fold cross-validation ini membantu menilai generalisasi model pada data baru dengan memperhitungkan variabilitas hasil antar fold dan mencegah overfitting pada subset tertentu dari data. Hasil prediksi ini memungkinkan evaluasi yang lebih akurat terhadap performa model dalam mendeteksi penipuan kartu kredit, dengan memanfaatkan seluruh data secara efektif untuk pelatihan dan pengujian secara bergantian.

E. Hasil Klasifikasi

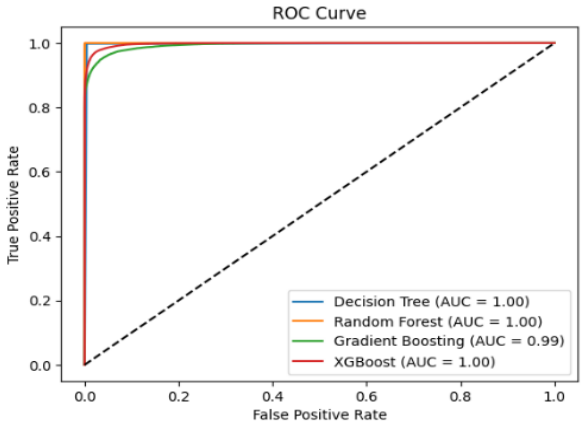
Pada tahapan ini dibandingkan dua metode preprocessing dan hasilnya disajikan pada Gambar 9,10 dan 11. Gambar 9 menyajikan hasil input preprocessing tahap pertama yang terdiri dari penghapusan nilai yang hilang, penghapusan duplikat data dan normalisasi proses. Gambar 10 menyajikan hasil implementasi oversampling smote-tomek pada model dan Gambar 11 menyajikan hasil implementasi oversampling smote-tomek yang dikombinasi dengan feature engineering dan cross validation. Hasilnya adalah semua metode yang digunakan memiliki kinerja terbaik disemua matriksnya berdasarkan analisis presisi, recall, f1-score dan ROC Curve (AUC). Penggunaan feature enginerring dan cross validation setelah penerapan SMOTE-Tomek berkontribusi positif dalam meningkatkan kinerja DT, RF, GB dan XGBoost, yang mana nilai yang dihasilkan pada setiap model hampir memiliki nilai yang sama, artinya feature engineering dan cross validation bekerja secara maksimal dalam menyeleksi fitur yang paling signifikan untuk meningkatkan kualitas data dan evaluasi yang lebih akurat terhadap performa model dalam mendeteksi penipuan kartu kredit.



Gambar 7. Hasil Model Tanpa Smote-Tomek



Gambar 8. Hasil Model Setelah Smote-Tomek



Gambar 9. Hasil Model Setelah Smote-Tomek, Feature Engineering & Cross Validation

Selanjutnya, dengan menambahkan teknik feature engineering dan cross validation setelah dilakukan inputan pada tahapan kedua yaitu teknik oversampling SMOTE-Tomek. Hasil menunjukan ada perubahan yang sangat signifikan dari setiap matriks dan berdasarkan analisis ROC Curve (AUC) hasil analisis disetiap metode hampir memiliki nilai seimbang yaitu (1.00 dan 0.99), seperti yang terlihat pada Gambar 11.

Pada percobaan di hasil Gambar 11 dengan menambahkan teknik feature engineering dan cross validation setelah dilakukan teknik oversampling Smote-Tomek, semua model pembelajaran mesin memiliki nilai yang sama rata tinggi disetiap matriksnya yaitu presisi, recall specifity dan F1 Score. Hasil ini memiliki nilai yang paling baik dibandingkan dengan hasil pada percobaan sebelumnya, yang mana performa dari setiap models belum bekerja secara maskimal sehingga masih ada nilai yang rendah pada matriks tertentu terutama pada presisi, recall dan f1-Score di setiap model yang digunakan. Tahapan percobaan tersebut menunjukan bahwa penggunaan feature engineering, cross validation dan SMOTE-Tomek memiliki kontribusi yang besar dalam peningkatan perfoma model di semua matriks yang ada.

TABEL I.
PERFORMA MODEL PEMBELAJARAN MESIN PADA DATASET ORIGINAL

Model	Presisi	Recall	Specificity	F1-Score	ROC AUC
Decision Tree	70.13%	77.69%	99.94%	0.73	0.89
Random Forest	**92.10%**	**80.76%**	99.98%	**0.86**	**0.96**
Gradient Boosting	89.10%	69.23%	99.98%	0.77	0.83
XGBoost	**91.81%**	77.69%	99.98%	0.84	**0.98**

TABEL II.
PERFORMA MODEL PEMBELAJARAN MESIN MENGGUNAKAN SMOTE-TOMEK TANPA FEATURE ENGINEERING

Model	Presisi	Recall	Specificity	F1-Score	ROC AUC
Decision Tree	32.91%	80.00%	99.75%	0.46	0.90
Random Forest	**84.37%**	83.07%	99.97%	**0.83**	**0.98**
Gradient Boosting	7.16%	**91.53%**	98.19%	0.13	**0.98**
XGBoost	71.89%	84.61%	99.94%	0.77	**0.98**

TABEL III.
PERFORMA MODEL PEMBELAJARAN MESIN MENGGUNAKAN SMOTE-TOMEK, FEATURE ENGINEERING & CROSS VALIDATION

Model	Presisi	Recall	Specificity	F1-Score	ROC AUC
Decision Tree	**99.55%**	99.78%	99.55%	**0.99**	**1.00**
Random Forest	**99.91%**	**99.98%**	99.91%	**0.99**	**1.00**
Gradient Boosting	97.66%	92.99%	97.77%	0.95	0.99
XGBoost	98.11%	96.47%	98.14%	0.97	**1.00**

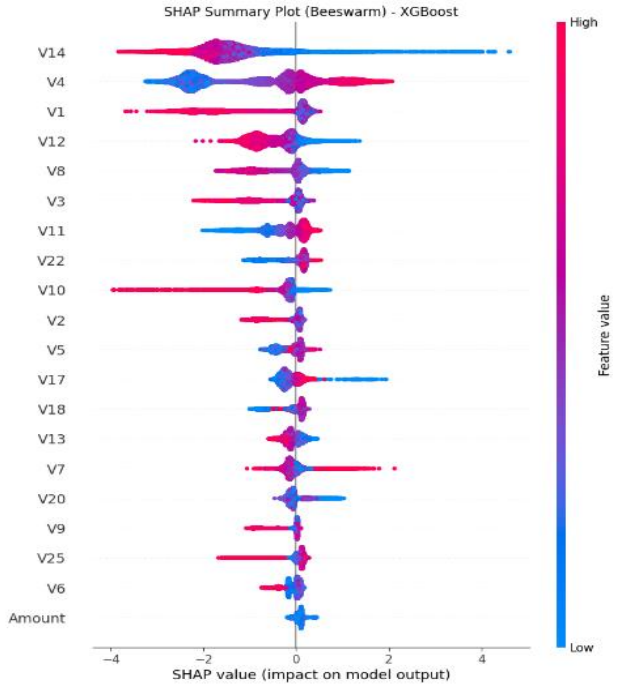
Hal ini menunjukkan efektivitas beberapa algoritma pembelajaran mesin, yaitu Random Forest, Decision Tree, Gradient Boosting dan XGBoost, dalam mendeteksi penipuan kartu kredit menggunakan teknik oversampling Smote-Tomek yang dikombinasikan dengan rekayasa fitur dan validasi silang. Hasilnya menunjukkan bahwa keempat model ini memberikan kinerja terbaik dalam mendeteksi transaksi penipuan, dengan peningkatan yang signifikan dalam recall dan precision setelah penerapan Smote-Tomek yang dikombinasikan dengan rekayasa fitur dan validasi silang. Menggunakan Smote-Tomek dalam kombinasi dengan rekayasa fitur dan validasi silang dapat meningkatkan keseimbangan data dan generalisasi model, meskipun meningkatkan waktu komputasi. Secara keseluruhan, penggunaan teknik dan model ensemble ini menunjukkan potensi besar untuk deteksi penipuan yang lebih akurat dan andal dalam industri keuangan. Penelitian di masa mendatang dapat mengeksplorasi algoritma yang lebih maju dan metode oversampling lainnya untuk lebih meningkatkan kinerja deteksi.

Melalui hasil pengujian pada Tabel 4 terlihat bahwa penerapan teknik oversampling menggunakan Smote-Tomek yang dikombinasikan dengan feature engineering dan cross validation secara signifikan meningkatkan performa model esemble learning untuk klasifikasi transaksi kecurigaan kartu kredit. Jika dilihat secara keseluruhan, semua model esemble learning memiliki hasil penilaian akurasi analisis yang sangat tinggi pada semua matriks, untuk detail nya bisa di lihat pada Tabel 3.

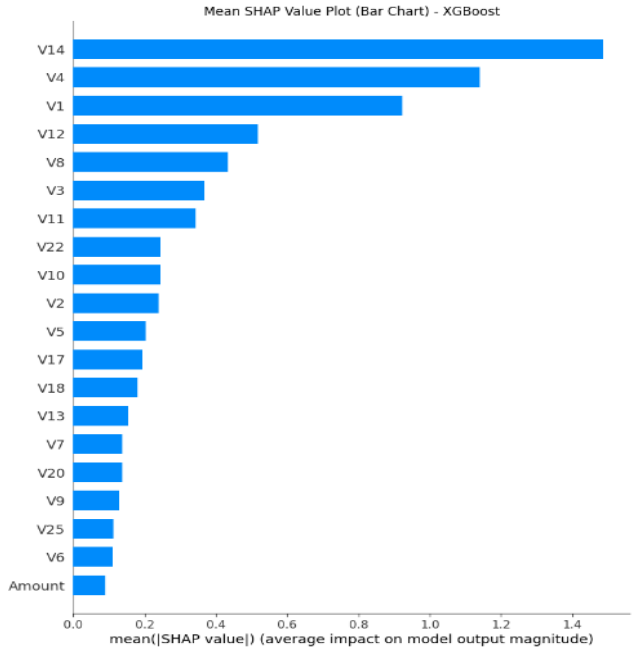
TABEL IV.
PERFORMA KOMPARASI DENGAN PENELITIAN LAIN

Reference	Alogaritma	Presisi	Recall	F1 Score
Aghware et al. 2024[17]	LR+Chi Squared	98.05	98.05	98.05
Aghware et al. 2024 [17]	KNN+ Chi Squared	90.18	94.48	92.10
Mrozek et al. 2020 [27]	Random forest-SMOTE	89.70	82.99	86.21
Al-Shabi 2019 [28]	Autoencoder (Thr=5)	0.64	0.011	0.19
Proposed Method	RF+Smote-Tomek, FE & CV	99.55	99.78	0.99
Proposed Method	DT+Smote-Tomek, FE & CV	99.91	99.98	0.99
Proposed Method	GB+Smote-Tomek, FE & CV	97.66	92.99	0.95
Proposed Method	XGB+Smote-Tomek, FE & CV	98.11	96.47	0.97

Hasil pengujian teknik oversampling Smote-Tomek yang dikombinasikan dengan feature engineering dan cross validation dapat meningkatkan kemampuan model untuk mengklasifikasi transaksi kecurigaan kartu kredit dengan lebih akurat dari pada model yang tidak menerapkan oversampling.



Gambar 10. Plot Beeswarm Menunjukan Feature Yang Paling Berdampak Pada Model XGBoost



Gambar 11. Feature Yang Terpenting Berdasarkan Nilai SHAP Absolute Rata-Rata Menggunakan XGBoost

Untuk meningkatkan interpretabilitas model pembelejaran mesin, kami mengintegrasikan nilai SHAP (Shapley Additive Explanations) untuk mendeskripsikan tentang kontribusi fitur individual terhadap prediksi model. SHAP memberikan pendekatan yang konsisten dan secara teoritis baik untuk kontribusi fitur dengan memastikan bahwa kontribusi fitur didistribusikan secara adil sesuai dengan kontribusi marginalnya di semua kemungkinan subset fitur [29]. Gambar 12 menyajikan plot SHAP beeswarm, yang menyediakan visualisasi terperinci tentang pentingnya dan dampak berbagai fitur dalam memprediksi penipuan kartu kredit menggunakan model XGBoost. Setiap titik dalam plot mewakili nilai SHAP untuk fitur tertentu, yang diberi kode warna berdasarkan nilai fitur (nilai tinggi berwarna warna merah dan nilai rendah berwarna biru). Jenis plot ini memberikan pemahaman yang lebih mendalam tentang pengaruh setiap fitur terhadap prediksi model. Misalnya, fitur V14 menunjukan nilai SHAP tertinggi, menunjukan bahwa fitur tersebut secara signifikan mempengaruhi prediksi model. Sebaliknya, V4 memiliki distribusi nilai tinggi dan rendahnya seimbang, yang menunjukan hubungan yang lebih kompleks dengan keluaran model. Kemampuan diagram sarang lebah untuk menunjukkan distribusi dan variabilitas dampak fitur, serta interaksi antara berbagai nilai fitur. Kemudian, pada Gambar 13 menunjukan diagram batang yang memberikan peringkat fitur bedasarkan nilai SHAP absolut rata-ratanya. Diagram batang memberikan peringkat yang jelas dan lugas tentang pentingnya fitur, diagram sarang lebah menawarkan wawasan yang lebih mendalam tentang efek kontekstual dan interaksi fitur pada prediksi model.

IV. KESIMPULAN

Penelitian ini menunjukan bahwa penerapan teknik penyeimbangan data Smote-Tomek yang dikombinasikan dengan feature engineering dan cross validation dapat meningkatkan performa model pembelajaran mesin dalam klasifikasi transaksi kecurigaan kartu kredit. Model Random Forest memiliki performa terbaik berdasarkan analisis ROC Curve dengan skor 1.00 dengan nilai presisi 99.91%, recall 99.98%, sepecificity 99.91% dan f1-score 0.99 dikarenakan model ini memiliki kemampuan untuk mengatasi overfitting lebih baik daripada Decision Tree. Ini dikarenakan hasil prediksi didasarkan pada rata-rata prediksi dari banyak pohon, yang menghasilkan model yang lebih stabil dan robust. Selain itu, dibandingkan dengan model lain yaitu Gradient Boosting yang sensitif terhadap noise, Random Forest lebih tahan terhadap variasi dan ketidakseimbangan data. Untuk XGBoost, meski memiliki kemampuan tinggi dalam akurasi, namun mungkin tidak optimal untuk data yang sangat tidak seimbang jika tidak ada tuning yang tepat.

Model Random Forest menunjukkan peningkatan kinerja yang signifikan setelah menerapkan teknik tersebut. Oleh karena itu, teknik oversampling yang di kombinasikan dengan feature engineering dan cross validation dapat diterapkan pada data credit card yang tidak seimbang untuk meningkatkan performa model klasifikasi.

DAFTAR PUSTAKA

[1] M. Habibpour *et al.*, "Uncertainty-aware credit card fraud detection using deep learning," *Eng. Appl. Artif. Intell.*, vol. 123, p. 106248,

2023, doi: 10.1016/j.engappai.2023.106248.
[2] G. Zhang *et al.*, "eFraudCom: An E-commerce Fraud Detection System via Competitive Graph Neural Networks," *ACM Trans. Inf. Syst.*, vol. 40, no. 3, pp. 1–29, Jul. 2022, doi: 10.1145/3474379.
[3] A. Ali *et al.*, "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review," *Appl. Sci.*, vol. 12, no. 19, p. 9637, Sep. 2022, doi: 10.3390/app12199637.
[4] P. T. S. Ningsih, M. Gusvarizon, and R. Hermawan, "Analisis Sistem Pendeteksi Penipuan Transaksi Kartu Kredit dengan Algoritma Machine Learning," *J. Teknol. Inform. dan Komput.*, vol. 8, no. 2, pp. 386–401, Sep. 2022, doi: 10.37012/jtik.v8i2.1306.
[5] A. Shen, R. Tong, and Y. Deng, "Application of Classification Models on Credit Card Fraud Detection," in *2007 International Conference on Service Systems and Service Management*, Jun. 2007, pp. 1–4. doi: 10.1109/ICSSSM.2007.4280163.
[6] E. F. Malik, K. W. Khaw, B. Belaton, W. P. Wong, and X. Chew, "Credit Card Fraud Detection Using a New Hybrid Machine Learning Architecture," *Mathematics*, vol. 10, no. 9, p. 1480, Apr. 2022, doi: 10.3390/math10091480.
[7] Y. Pristyanto and A. A. Zein, "Model Balanced Bagging Berbasis Decision Tree Pada Dataset Imbalanced Class," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 9–15, 2023, doi: 10.32736/sisfokom.v12i1.1399.
[8] E. Ileberi, Y. Sun, and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," *IEEE Access*, vol. 9, pp. 165286–165294, 2021, doi: 10.1109/access.2021.3134330.
[9] M. Mujahid *et al.*, "Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00943-4.
[10] D. R. I. M. Setiadi, H. M. M. Islam, G. A. Trisnapradika, and W. Herowati, "Analyzing Preprocessing Impact on Machine Learning Classifiers for Cryotherapy and Immunotherapy Dataset," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 39–50, Jun. 2024, doi: 10.62411/faith.2024-2.
[11] Z. S. Dhahir, "A Hybrid Approach for Efficient DDoS Detection in Network Traffic Using CBLOF-Based Feature Engineering and XGBoost," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 174–190, Sep. 2024, doi: 10.62411/faith.2024-33.
[12] F. Omoruwou, A. A. Ojugo, and S. E. Ilodigwe, "Strategic Feature Selection for Enhanced Scorch Prediction in Flexible Polyurethane Form Manufacturing," *J. Comput. Theor. Appl.*, vol. 1, no. 3, pp. 346–357, Feb. 2024, doi: 10.62411/jcta.9539.
[13] J. A. Ingio, A. S. Nsang, and A. Iorliam, "Optimizing Rice Production Forecasting Through Integrating Multiple Linear Regression with Recursive Feature Elimination," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 96–108, Aug. 2024, doi: 10.62411/faith.2024-17.
[14] M. I. Akazue, I. A. Debekeme, A. E. Edje, C. Asuai, and U. J. Osame, "UNMASKING FRAUDSTERS: Ensemble Features Selection to Enhance Random Forest Fraud Detection," *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 201–211, Dec. 2023, doi: 10.33633/jcta.v1i2.9462.
[15] D. R. I. M. Setiadi, S. Widiono, A. N. Safriandono, and S. Budi, "Phishing Website Detection Using Bidirectional Gated Recurrent Unit Model and Feature Selection," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 1, pp. 75–83, 2024, doi: 10.62411/faith.2024-15.
[16] M. D. Okpor *et al.*, "Pilot Study on Enhanced Detection of Cues over Malicious Sites Using Data Balancing on the Random Forest Ensemble," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 109–123, Sep. 2024, doi: 10.62411/faith.2024-14.
[17] F. O. Aghware *et al.*, "Enhancing the Random Forest Model via Synthetic Minority Oversampling Technique for Credit-Card Fraud Detection," *J. Comput. Theor. Appl.*, vol. 1, no. 4, pp. 407–420, Mar. 2024, doi: 10.62411/jcta.10323.
[18] Machine Learning Group - ULB, "Credit Card Fraud Detection," *Kaggle.com*, 2017. https://kaggle.com/mlg-ulb/creditcardfraud
[19] F. . Osisanwo, J. E. . Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised Machine Learning Algorithms: Classification and Comparison," *Int. J. Comput. Trends Technol.*, vol. 48, no. 3, pp. 128–138, 2017, doi: 10.14445/22312803/ijctt-v48p126.
[20] Y. Xin and X. Ren, "Predicting depression among rural and urban disabled elderly in China using a random forest classifier," *BMC Psychiatry*, vol. 22, no. 1, p. 118, Feb. 2022, doi: 10.1186/s12888-022-03742-4.
[21] X. Y. Liew, N. Hameed, and J. Clos, "An investigation of XGBoost-based algorithm for breast cancer classification," *Mach. Learn. with Appl.*, vol. 6, p. 100154, Dec. 2021, doi: 10.1016/j.mlwa.2021.100154.
[22] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *J. Chemom.*, vol. 18, no. 6, pp. 275–285, 2004, doi: 10.1002/cem.873.
[23] O. Lyashevska, F. Malone, E. MacCarthy, J. Fiehler, J.-H. Buhk, and L. Morris, "Class imbalance in gradient boosting classification algorithms: Application to experimental stroke data," *Stat. Methods Med. Res.*, vol. 30, no. 3, pp. 916–925, 2020, doi: 10.1177/0962280220980484.
[24] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A Review," *Int. J. Adv. Soft Compu. Appl*, vol. 7, no. 3, pp. 176–204, 2015.
[25] G. Lemaitre, F. Nogueira, and C. K. Aridas, "Imbalanced-Learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 559–563, Jan. 2017.
[26] M. Kubat, "Addressing the Curse of Imbalanced Training Sets: One-Sided Selection," *Fourteenth Int. Conf. Mach. Learn.*, 2000.
[27] P. Mrozek, J. Panneerselvam, and O. Bagdasar, "Efficient Resampling for Fraud Detection During Anonymised Credit Card Transactions with Unbalanced Datasets," *2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing (UCC)*. IEEE, 2020. doi: 10.1109/ucc48980.2020.00067.
[28] M. A. Al-Shabi, "Credit Card Fraud Detection Using Autoencoder Model in Unbalanced Datasets," *J. Adv. Math. Comput. Sci.*, pp. 1–16, 2019, doi: 10.9734/jamcs/2019/v33i530192.
[29] T. R. Noviandy, G. M. Idroes, and I. Hardi, "An Interpretable Machine Learning Strategy for Antimalarial Drug Discovery with LightGBM and SHAP," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 84–95, Aug. 2024, doi: 10.62411/faith.2024-16.