

**Penerapan Extreme Gradient Boosting (XGBoost)
dengan SMOTE untuk Deteksi Penipuan Kartu Kredit**

TUGAS AKHIR

**Nicholas Anthony Suhartono
1118049**



**PROGRAM STUDI INFORMATIKA
INSTITUT TEKNOLOGI HARAPAN BANGSA
BANDUNG
2022**

**Penerapan Extreme Gradient Boosting (XGBoost)
dengan SMOTE untuk Deteksi Penipuan Kartu Kredit**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh
gelar sarjana dalam bidang Informatika**

**Nicholas Anthony Suhartono
1118049**



**PROGRAM STUDI INFORMATIKA
INSTITUT TEKNOLOGI HARAPAN BANGSA
BANDUNG
2022**

ABSTRAK

Nama : Nicholas Anthony Suhartono
Program : Informatika
Studi
Judul : Penerapan Extreme Gradient Boosting (XGBoost)
dengan SMOTE untuk Deteksi Penipuan Kartu Kredit

Kartu kredit merupakan metode pembayaran yang umum digunakan oleh masyarakat. Kejahatan Penipuan kartu kredit adalah bentuk kejahatan penipuan yang melibatkan metode pembayaran kartu kredit yang terjadi ketika pelaku kejahatan menggunakan data pemilik kartu kredit untuk melakukan transaksi. Kejahatan ini mengalami peningkatan yang signifikan setiap tahunnya dan mengakibatkan kerugian yang cukup besar. Dataset penipuan kartu kredit memiliki jumlah data antar kelas yang sangat tidak seimbang (*imbalance*) sehingga menjadi salah satu masalah untuk melakukan pelatihan pada model *machine learning*. Pada penelitian ini, akan dilakukan klasifikasi menggunakan *Extreme Gradient Boosting* (XGBoost). Untuk mengatasi masalah data latih yang terbatas, akan dilakukan *oversampling* menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE). Proses pelatihan akan menggunakan metode *k-fold cross validation* serta *grid search* untuk mencari *hyperparameter* terbaik yang cukup sensitif untuk mendeteksi penipuan kartu kredit. Untuk mendapatkan nilai sensitivitas tertinggi, maka pengujian akan menggunakan *confusion matrix* sebagai pengukuran khususnya untuk mencari nilai *recall* tertinggi. Berdasarkan hasil pengujian yang dilakukan, nilai *recall* tertinggi yang diperoleh sebesar 85.26% yang diperoleh dengan menggunakan *hyperparameter* $n_estimators = 50$, $max_depth = 5$, $eta = 0.3$, $lambda = 0.5$, $gamma = 0.5$.

Kata kunci: Klasifikasi, *Oversampling*, Penipuan Kartu Kredit, *Imbalance*, *Extreme Gradient Boosting* (XGBoost), *Synthetic Minority Oversampling Technique* (SMOTE).

ABSTRACT

Name : Nicholas Anthony Suhartono
Department : Informatics
Title : Implementation of Extreme Gradient Boosting (XGBoost)
and SMOTE for Credit Card Fraud Detection.

Credit card is one of many payment methods that are commonly used globally. Credit card fraud is a fraud that involves stealing data from credit card owners and using it to make transactions. This type of fraud has significantly increased and resulted in substantial financial losses. The credit card fraud dataset typically includes imbalance classes because data labelled as fraud rarely occurs within normal transactions and becomes a challenge for machine learning classification algorithms. In this research, Extreme Gradient Boosting (XGBoost) will be implemented for classification along with Synthetic Minority Oversampling Technique (SMOTE) to oversample minority classes. K-fold Cross-validation and Grid Search are implemented to find the best XGBoost hyperparameter which is sensitive enough to detect a fraud transaction. Based on the test results, the highest recall that XGBoost achieved is 85.26% when hyperparameter $n_estimators = 50$, $max_depth = 5$, $eta = 0.3$, $lambda = 0.5$, $gamma = 0.5$.

Keywords: Classification, Oversampling, Credit Card Fraud, Imbalance, Extreme Gradient Boosting (XGBoost), Synthetic Minority Oversampling Technique (SMOTE).

KATA PENGANTAR

Penulis mengucapkan terima kasih kepada Tuhan Yesus karena dengan karunia-Nya, penulis dapat menyelesaikan Tugas Akhir berjudul Penerapan Extreme Gradient Boosting (XGBoost) dengan SMOTE untuk Deteksi Penipuan Kartu Kredit. Laporan ini disusun sebagai salah satu syarat kelulusan di Institut Teknologi Harapan Bangsa. Pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Tuhan Yesus Kristus, karena oleh bimbingan-Nya penulis selalu mendapat pengharapan untuk menyelesaikan tugas akhir ini.
2. Bapak Ventje J. Lewi Engel, M.T., CEH, selaku pembimbing utama Tugas Akhir yang senantiasa memberikan dukungan, semangat, dan meluangkan waktu bimbingan yang memberikan ilmu, wawasan, dan saran kepada penulis.
3. Bapak Sriwisnu Noloadi, S.Sc., M.Kom., selaku penguji I Tugas Akhir yang telah memberikan saran dan masukan-masukan kritis kepada penulis.
4. Bapak Yoyok Yusman Gamaliel, S.T.,M. Eng., selaku penguji II Tugas Akhir yang telah memberikan saran dan masukan-masukan kritis kepada penulis.
5. Seluruh dosen dan staff Departemen Teknik Informatika ITHB yang telah membantu dalam menyelesaikan Laporan Tugas Akhir ini.
6. Orang tua penulis yang selalu meluangkan waktu dan memberikan doa, semangat, nasihat, dan dukungan kepada penulis untuk menyelesaikan Laporan Tugas Akhir ini.
7. Evelyn Melissa Suhartono, S.Hub.Int., selaku kakak penulis yang telah membantu dalam visualisasi dan desain.
8. Teman seangkatan, adik dan kakak kelas yang berkuliah di satu tempat dan berbeda tempat, yang telah memberikan dukungan, bantuan teknis, semangat, dan waktu diskusi kepada penulis selama melakukan Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari sempurna, baik dari segi konten maupun penulisan, karena keterbatasan waktu, pengetahuan, dan pengalaman yang dimiliki oleh penulis. Penulis menyampaikan permohonan maaf atas hal tersebut. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi penyempurnaan penelitian sejenis di masa mendatang.

Bandung, 5 Juli 2022

Hormat penulis,

A handwritten signature in black ink, appearing to read 'Nicholas Anthony Suhartono', with a long horizontal stroke extending from the bottom right.

Nicholas Anthony Suhartono

DAFTAR ISI

ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	xii
DAFTAR GAMBAR	xv
DAFTAR ALGORITMA	xvi
DAFTAR LAMPIRAN	xvii
BAB 1 PENDAHULUAN	1-1
1.1 Latar Belakang	1-1
1.2 Rumusan Masalah	1-3
1.3 Tujuan Penelitian	1-3
1.4 Batasan Masalah	1-4
1.5 Kontribusi Penelitian	1-4
1.6 Metodologi Penelitian	1-4
1.7 Sistematika Pembahasan	1-5
BAB 2 LANDASAN TEORI	2-1
2.1 Tinjauan Pustaka	2-1
2.1.1 <i>Decision Tree</i>	2-1
2.1.2 <i>Ensemble Learning</i>	2-2
2.1.3 <i>Random Forest</i>	2-3
2.1.4 <i>Gradient Boost</i>	2-6
2.1.4.1 <i>XGBoost</i>	2-9
2.1.4.2 <i>Loss Function</i>	2-14
2.1.4.3 <i>Hyperparameter</i> dalam XGBoost	2-14
2.1.5 Optimisasi pada <i>Library</i> XGBoost	2-15
2.1.5.1 Algoritme <i>Basic Exact Greedy</i>	2-16
2.1.5.2 Algoritme <i>Approximate split-finding</i>	2-16

2.1.5.3	Sparsity-aware Split Finding	.2-16
2.1.5.4	<i>Parallel Learning</i>	.2-18
2.1.5.5	Cache-aware Access	.2-18
2.1.5.6	<i>Blocks for Out-of-core Computation</i>	.2-19
2.1.6	<i>Korelasi</i>	.2-19
2.1.6.1	<i>Pearson Correlation Coefficient</i>	.2-19
2.1.6.2	Spearman Rank Correlation Coefficient	.2-20
2.1.7	<i>Resampling</i>	.2-21
2.1.8	<i>SMOTE</i>	.2-22
2.1.9	<i>Hyperparameter Tuning</i>	.2-29
2.1.10	<i>K-Fold Cross-Validation</i>	.2-30
2.1.11	<i>Confusion Matrix</i>	.2-31
2.1.12	Pustaka Python	.2-33
2.1.12.1	Pandas	.2-33
2.1.12.2	Numpy	.2-34
2.1.12.3	Scikit-Learn	.2-36
2.1.12.4	Imbalanced-Learn	.2-36
2.1.12.5	Matplotlib	.2-37
2.1.12.6	Seaborn	.2-38
2.1.12.7	XGboost	.2-38
2.2	Tinjauan Studi	.2-39
2.2.1	<i>State Of The Art</i>	.2-39
2.3	Tinjauan Objek	.2-41
2.3.1	Kartu Kredit	.2-41
2.3.2	Standar Keamanan Data dalam Industri Kartu Pembayaran	.2-41
2.3.2.1	Transaksi Pembayaran Kartu Kredit	.2-42
2.3.2.2	Penipuan Kartu Kredit	.2-44
2.3.3	<i>Dataset</i> Transaksi Kartu Kredit	.2-44

BAB 3 ANALISIS DAN PERANCANGAN SISTEM 3-1

3.1	Analisis Masalah	3-1
3.2	Kerangka Pemikiran	3-1
3.2.1	Penjelasan Indikator	3-2
3.3	Analisis Urutan Proses Global	3-3
3.4	Analisis Data Sampling	3-6
3.5	Data <i>Preprocessing</i>	3-6
3.6	Data <i>splitting</i>	3-9

3.7	<i>Oversampling</i> dengan SMOTE	3-9
3.8	Analisis kasus XGBoost	3-10
3.8.1	Pembentukan <i>Tree</i> Pertama	3-11
3.8.2	Pembentukan <i>Tree</i> Kedua	3-18
3.8.3	Melakukan prediksi	3-24
3.9	Analisis Manual Perhitungan Korelasi	3-26
3.9.1	Perhitungan Pearson Correlation Coefficient	3-27
3.9.2	Perhitungan Spearman Rank Correlation Coefficient	3-31

BAB 4 IMPLEMENTASI DAN PENGUJIAN **4-1**

4.1	Lingkungan Implementasi	4-1
4.1.1	Spesifikasi Perangkat Keras	4-1
4.1.2	Spesifikasi Perangkat Lunak	4-1
4.2	Implementasi Perangkat Lunak	4-1
4.2.1	<i>Class</i> Preprocessing	4-2
4.2.2	<i>Class</i> Node	4-3
4.2.3	<i>Class</i> XGBTree	4-4
4.2.4	<i>Class</i> XGBClassifierBase	4-4
4.2.5	<i>Class</i> XGBClassifier	4-5
4.2.6	<i>Class</i> GridSearch	4-5
4.2.7	CrossValidation	4-6
4.2.8	<i>Class</i> Evaluation	4-6
4.3	Pengujian	4-7
4.3.1	Skenario Pengujian XGBoost	4-8
4.3.2	Pengujian XGBoost yang Dibangun Dari <i>Scratch</i>	4-11
4.3.3	Pengujian XGBoost menggunakan <i>5-fold</i> <i>Cross-Validation</i> dan <i>Grid Search</i>	4-23
4.3.3.1	Pengujian Estimators 50	4-23
4.3.3.2	Pengujian Estimators 100	4-26
4.3.3.3	Pengujian Estimators 200	4-30
4.3.3.4	Pengujian XGBoost dengan Seleksi Fitur	4-33
4.3.4	Pembahasan Pengujian	4-35
4.3.5	Pengujian XGBoost pada Data Uji	4-41
4.3.5.1	Pengujian XGBoost pada Data Uji tanpa SMOTE	4-45
4.3.6	Pengujian XGBoost pada Data yang Berbeda	4-47

BAB 5 KESIMPULAN DAN SARAN **5-1**

5.1	Kesimpulan	5-1
5.2	Saran	5-3
Daftar Referensi		i
BAB A		
	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-1
BAB B		
	Hasil Rata-rata Pengujian XGBoost dengan SMOTE menggunakan Cross Validation	B-15

DAFTAR TABEL

2.1	Daftar masukan SMOTE [22]	2-27
2.2	Daftar parameter SMOTE [22]	2-28
2.3	Daftar keluaran SMOTE [22]	2-29
2.4	Daftar Metode yang Digunakan	2-33
2.5	Daftar Metode yang Digunakan	2-34
2.6	Daftar Metode yang Digunakan	2-36
2.7	Daftar Metode yang Digunakan	2-36
2.8	Daftar Metode yang Digunakan	2-37
2.9	Daftar Metode yang Digunakan	2-38
2.10	Daftar Metode yang Digunakan	2-38
2.11	Tinjauan Studi	2-39
3.1	Dataset untuk analisis kasus.	3-10
3.2	Inisialisasi nilai prediksi dan probabilitas untuk setiap data latih	3-11
3.3	Menghitung nilai <i>gradient</i> dan <i>hessian</i> untuk setiap data latih.	3-11
3.4	Data pada fitur v1 yang telah diurutkan.	3-13
3.5	Hasil perhitungan <i>gradient</i> , <i>hessian</i> , dan <i>gain</i> pada <i>tree</i> pertama fitur v1.	3-14
3.6	Hasil perhitungan <i>gradient</i> , <i>hessian</i> , dan <i>gain tree</i> pertama fitur v2.	3-15
3.7	Tabel pada <i>leaf node</i> kiri berdasarkan gambar 3.7	3-16
3.8	Tabel pada <i>leaf node</i> kanan berdasarkan gambar 3.7	3-16
3.9	Hasil prediksi dan probabilitas baru untuk setiap data latih . .	3-19
3.10	Nilai <i>gradient</i> dan <i>hessian</i> untuk <i>tree</i> kedua	3-19
3.11	Nilai <i>gradient</i> , <i>hessian</i> , dan <i>gain tree</i> kedua fitur v1.	3-20
3.12	Hasil perhitungan <i>gradient</i> , <i>hessian</i> , dan <i>gain tree</i> kedua fitur v2.	3-21
3.13	Hasil prediksi akhir	3-24
3.14	Dataset yang digunakan untuk analisis manual korelasi	3-26
3.15	Hasil perhitungan $x - \bar{x}$ dan $y - \bar{y}$ menggunakan metode <i>pearson</i>	3-27
3.16	Hasil perhitungan $(x - \bar{x})^2$ dan $(y - \bar{y})^2$ menggunakan metode <i>pearson</i>	3-28
3.17	Hasil perhitungan $(x - \bar{x})(y - \bar{y})$ menggunakan metode <i>pearson</i>	3-29
3.18	Hasil <i>ranking</i> x dan y	3-31
3.19	Hasil <i>ranking</i> menjadi data x dan y	3-32

3.20	Hasil perhitungan $x - \bar{x}$ dan $y - \bar{y}$ pada metode <i>spearman</i> . . .	3-33
3.21	Hasil perhitungan $(x - \bar{x})^2$ dan $(y - \bar{y})^2$ pada metode <i>spearman</i>	3-34
3.22	Hasil perhitungan $(x - \bar{x})(y - \bar{y})$ pada metode <i>spearman</i> . . .	3-35
4.1	Daftar Metode <i>Class</i> Preprocessing	4-2
4.2	Daftar Metode <i>Class</i> Node	4-3
4.3	Daftar Metode <i>Class</i> XGBTree	4-4
4.4	Daftar Metode <i>Class</i> XGBClassifierBase	4-4
4.5	Daftar Metode <i>Class</i> XGBClassifier	4-5
4.6	Daftar Metode <i>Class</i> GridSearch	4-6
4.7	Daftar Metode <i>Class</i> CrossValidation	4-6
4.8	Daftar Metode <i>Class</i> Evaluation	4-6
4.9	Total kombinasi kasus pengujian XGBoost.	4-8
4.10	Daftar kombinasi kasus pengujian XGBoost	4-8
4.11	Hasil Pengujian XGBoost dengan Estimators 50	4-24
4.12	Hasil Pengujian XGBoost dengan Estimators 100	4-27
4.13	Hasil Pengujian XGBoost dengan Estimators 200	4-31
4.14	Kombinasi hyperparameter yang menghasilkan nilai <i>recall</i> , <i>f-score</i> , dan akurasi tertinggi	4-35
4.15	Hasil Pengujian XGBoost pada Data Uji	4-42
4.16	Hasil Pengujian XGBoost pada data uji tanpa SMOTE . . .	4-45
4.17	Hasil pengujian XGBoost pada <i>dataset</i> yang ditukar	4-48
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-1
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-2
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-3
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-4
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-5
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-6
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-7

A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-8
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-9
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-10
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-11
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-12
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-13
A-1	Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-14
B-1	<i>Hasil Rata-rata pengujian XGBoost dengan SMOTE menggunakan Cross Validation</i>	B-15
B-1	<i>Hasil Rata-rata pengujian XGBoost dengan SMOTE menggunakan Cross Validation</i>	B-16
B-1	<i>Hasil Rata-rata pengujian XGBoost dengan SMOTE menggunakan Cross Validation</i>	B-17

DAFTAR GAMBAR

2.1	Ilustrasi <i>Decision Tree</i>	2-2
2.2	Ilustrasi alur proses pelatihan pada <i>bagging</i> dan <i>boosting</i>	2-3
2.3	Ilustrasi <i>Random Forest</i> ketika melakukan prediksi.	2-5
2.4	Ilustrasi Algoritme Gradient Boosting.	2-6
2.5	Struktur <i>tree</i> dengan <i>default direction</i> [3].	2-18
2.6	Ilustrasi struktur blok untuk <i>parallel learning</i> [3].	2-18
2.7	Jenis-jenis korelasi	2-20
2.8	Ilustrasi dataset yang <i>imbalance</i>	2-22
2.9	Ilustrasi <i>resampling</i>	2-22
2.10	Ilustrasi <i>oversampling</i> menggunakan SMOTE	2-27
2.11	Perbandingan proporsi jumlah data pada suatu kelas menggunakan <i>SMOTE</i> [22]	2-27
2.12	<i>k-fold cross-validation</i> dengan <i>fold</i> sebanyak 5 [16]	2-31
2.13	Ilustrasi <i>Confusion Matrix</i> [13].	2-32
2.14	PCI DSS	2-41
2.15	Ilustrasi proses transaksi pembayaran menggunakan kartu kredit pada jaringan <i>Mastercard</i>	2-44
2.16	Kelas yang tidak seimbang pada <i>dataset</i> yang akan digunakan. Kelas 1 merupakan kelas <i>fraud</i> atau transaksi penipuan dan kelas 0 merupakan kelas <i>non-fraud</i> atau transaksi bukan penipuan.	2-45
3.1	Kerangka Pemikiran	3-2
3.2	<i>Flowchart</i> Proses Global	3-4
3.3	Dataset transaksi pembayaran kartu kredit	3-6
3.4	<i>Pearson Heatmap Correlation Matrix</i>	3-7
3.5	<i>Spearman Heatmap Correlation Matrix</i>	3-8
3.6	Data latih setelah dilakukan <i>oversampling</i> dengan SMOTE pada fold ke 1.	3-10
3.7	Visualisasi <i>tree</i> pertama setelah melakukan <i>split</i>	3-16
3.8	Visualisasi penelusuran <i>tree</i> pertama.	3-18
3.9	Visualisasi <i>tree</i> kedua setelah melakukan <i>split</i>	3-22
3.10	Visualisasi penelusuran <i>tree</i> kedua.	3-23
4.1	Perbandingan waktu pelatihan	4-21
4.2	Hasil Pengujian pada Estimator 50	4-25

4.3	Hasil pengujian yang telah diurutkan dari nilai tertinggi ke nilai terendah pada <i>estimators</i> 50	.4-25
4.4	Hasil Pengujian pada Estimator 100	.4-28
4.5	Hasil pengujian yang telah diurutkan dari nilai tertinggi ke nilai terendah pada <i>estimators</i> 100	.4-29
4.6	Hasil Pengujian pada Estimator 200	.4-32
4.7	Hasil pengujian yang telah diurutkan dari nilai tertinggi ke nilai terendah pada <i>estimators</i> 200	.4-33
4.8	Hasil Pengujian dengan Seleksi Fitur	.4-34
4.9	Matriks Korelasi Pengukuran Pengujian dengan <i>Hyperparameter</i> yang diuji.	.4-36
4.10	Hasil Pengujian setiap <i>estimator</i>	.4-37
4.11	Hasil pengukuran <i>n_estimators</i> pada kombinasi pengujian pada nilai <i>hyperparameter</i> yang sama terkecuali <i>n_estimators</i>	.4-38
4.12	Hasil Pengujian <i>max_depth</i>	.4-40
4.13	Hasil Pengujian <i>lambda</i> dan <i>gamma</i>	.4-41
4.14	Beberapa fitur utama yang dikumpulkan dalam transaksi kartu kredit [29].	.4-43
4.15	Contoh perhitungan rata-rata waktu terjadinya transaksi kartu kredit yang kurang tepat [29].	.4-44
4.16	Perbandingan matriks korelasi tanpa SMOTE dengan menggunakan SMOTE	.4-46
4.17	Perbandingan matriks korelasi pengukuran pengujian dengan <i>hyperparameter</i> yang diuji tanpa SMOTE dan dengan menggunakan SMOTE	.4-46
4.18	Perbandingan proporsi kelas pada data latih dan data uji . . .	.4-48
4.19	Proporsi kelas pada data latih setelah dilakukan <i>oversampling</i> SMOTE	.4-48
4.20	Perbandingan penyebaran data sebelum dan setelah dilakukan <i>oversampling</i> SMOTE	.4-49

DAFTAR ALGORITMA

2.1	Algoritme <i>Gradient Boost</i> [1]	2-6
2.2	Algoritme XGBoost [2]	2-10
2.3	Algoritme <i>exact greedy</i> [3]	2-16
2.4	Algoritme <i>approximate split-finding</i> [3]	2-17
2.5	Algoritme <i>Sparsity-aware Split Finding</i>	2-17
2.6	Algoritme SMOTE	2-24
2.7	Algoritme <i>populate</i> pada SMOTE	2-25

DAFTAR LAMPIRAN

Hasil Pengujian XGBoost dengan SMOTE menggunakan Cross Validation Pada Setiap <i>Fold</i>	A-1
Hasil Rata-rata pengujian XGBoost dengan SMOTE menggunakan Cross Validation	B-15

DAFTAR REFERENSI

- [1] J. H. Friedman, “Greedy function approximation: A gradient boosting machine.”, *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001. DOI: 10.1214/aos/1013203451. [Online]. Available : <https://doi.org/10.1214/aos/1013203451>, [Accessed: 29 January 2022].
- [2] D.-K. Choi, “Data-driven materials modeling with xgboost algorithm and statistical inference analysis for prediction of fatigue strength of steels”, *International Journal of Precision Engineering and Manufacturing*, vol. 20, pp. 129–138, 2019. DOI: 10 . 1007 / s12541 - 019 - 00048 - 6. [Online]. Available : <https://doi.org/10.1007/s12541-019-00048-6>, [Accessed: 01 February 2022].
- [3] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system”, *CoRR*, vol. abs/1603.02754, 2016. arXiv: 1603.02754. [Online]. Available : <http://arxiv.org/abs/1603.02754>, [Accessed: 10 October 2021].
- [4] *Credit card fraud statistics*. [Online]. Available : <https://shiftprocessing.com/credit-card-fraud-statistics>, [Accessed: 5 December 2021].
- [5] S. P. Maniraj, A. Saini, S. Ahmed and S. D. Sarkar, “Credit card fraud detection using machine learning and data science”, *International Journal of Engineering Research Technology (IJERT)*, vol. 8, no. 9, 2019. DOI: <http://dx.doi.org/10.17577/IJERTV8IS090031>, [Accessed: 27 January 2022].
- [6] I. Kaur and M. Kalra, “Ensemble classification method for credit card fraud detection”, *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 3, pp. 25 579–25 587, 2019. DOI: 10.35940/ijrte.C4213.098319, [Accessed: 27 January 2022].
- [7] A. A. Taha and S. J. Malebary, “An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine”, *IEEE Access*, vol. 8, pp. 25 579–25 587, 2020. DOI: 10.1109/ACCESS.2020.2971354, [Accessed: 10 October 2021].
- [8] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid and H. Zeineddine, “An experimental study with imbalanced classification approaches for credit card fraud detection”, *IEEE Access*, vol. 7, pp. 93 010–93 022, 2019. DOI: 10 . 1109 / ACCESS . 2019 . 2927266, [Accessed: 10 October 2021].

- [9] S. Marabad, “Credit card fraud detection using machine learning”, *Asian Journal of Convergence in Technology*, vol. 7, no. 2, 2021, ISSN: 2350-1146. DOI: <https://doi.org/10.33130/AJCT.2021v07i02.023>, [Accessed: 27 January 2022].
- [10] *Credit card fraud detection*. [Online]. Available : <https://www.kaggle.com/mlg-ulb/creditcardfraud>, [Accessed: 10 October 2021].
- [11] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O’Reilly Media, Inc., 2019, ISBN: 9781492032649.
- [12] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*. Springer, Boston, MA, 2005, ISBN: 9780387254654.
- [13] A. Burkov, *The Hundred-Page Machine Learning Book*. Andriy Burkov, 2019, ISBN: 9781999579517.
- [14] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python*. O’Reilly Media, Inc., 2016, ISBN: 9781449369415.
- [15] A. Cutler, D. Cutler and J. Stevens, “Random forests”, in. Jan. 2011, vol. 45, pp. 157–176, ISBN: 978-1-4419-9325-0. DOI: 10.1007/978-1-4419-9326-7_5.
- [16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, “Scikit-learn: Machine learning in Python”, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available : <https://scikit-learn.org/>, [Accessed: 10 October 2021].
- [17] C. Wade and K. Glynn, *Hands-On Gradient Boosting with XGBoost and scikit-learn*. 2020, ISBN: 9781839218354.
- [18] *Fighting overfitting with l1 or l2 regularization: Which one is better?* [Online]. Available : <https://neptune.ai>, [Accessed: 12 February 2022].
- [19] *Xgboost documentation*. [Online]. Available : <https://xgboost.readthedocs.io>, [Accessed: 30 January 2022].
- [20] R. N. Forthofer, E. S. Lee and M. Hernandez, *Biostatistics: A Guide to Design, Analysis and Discovery*. Academic Press; 2007, ISBN: 9781449369415.
- [21] M. Swamynathan, *Mastering Machine Learning with Python in Six Steps*. 2017, ISBN: 9781484228661.

- [22] *Smote*. [Online]. Available : <https://docs.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/smote#expected-input>, [Accessed: 29 January 2022].
- [23] T. Agrawal, *Hyperparameter Optimization in Machine Learning*. 2021, ISBN: 9781484265796.
- [24] T. Ma, L. Wu, S. Zhu and H. Zhu, "Multiclassification prediction of clay sensitivity using extreme gradient boosting based on imbalanced dataset", *Applied Sciences*, vol. 12, no. 3, 2022, ISSN: 2076-3417. DOI: 10.3390/app12031143. [Online]. Available : <https://www.mdpi.com>.
- [25] *Credit card*. [Online]. Available : <https://www.investopedia.com/terms/c/creditcard.asp>, [Accessed: 21 February 2021].
- [26] *Pci dss quick reference guide: Understanding the payment card industry data security standard version 3.2.1*. [Online]. Available : <https://www.pcisecuritystandards.org/>, [Accessed: 03 March 2022].
- [27] *How the payment process works*. [Online]. Available : <https://sea.mastercard.com/en-region-sea/business/merchants/start-accepting/payment-process.html>, [Accessed: 21 February 2021].
- [28] D. K. K. Kajal, "Credit card fraud detection using imbalance resampling method with feature selection", *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, Jun. 2021. DOI: <https://doi.org/10.30534/ijatcse/2021/811032021>.
- [29] A. Correa Bahnsen, D. Aouada, A. Stojanovic and B. Ottersten, "Feature engineering strategies for credit card fraud detection", *Expert Systems with Applications*, vol. 51, Jan. 2016. DOI: 10.1016/j.eswa.2015.12.030.