



Penerapan Algoritma Extreme Gradient Boosting (Xgboost) Untuk Analisis Risiko Kredit

Aldo Andika Saputra¹, Betha Nurina Sari², Chaerur Rozikin³

^{1,2,3}Universitas Singaperbangsa Karawang

Received: 9 Desember 2023
Revised: 22 Desember 2023
Accepted: 30 Desember 2023

Abstrak

Based on data on outstanding online loans, in March this year non-current credit (30-90 days) jumped by IDR 3.58 trillion (71.13% YoY) with the highest individual loans amounting to IDR 3.3 trillion and the rest from business entities. Then those experiencing bad credit exceeding 90 days in March 2023 amounted to IDR 1.43 trillion of the total remaining loans outstanding by debtors. This value increased significantly by IDR 1.35 trillion (65.33%) compared to the previous period (YoY), where IDR 1.14 trillion was bad credit from individuals. This happens because it starts from a lack of accuracy in the customer screening process. This research aims to classify potential debtors using data mining techniques using the XGBoost algorithm. The method or steps to achieve the objectives of this research is the Knowledge Discovery in Database (KDD) methodology, which consists of five steps, namely data selection, data preprocessing, data transformation, data mining, and evaluation. The dataset consists of 1000 rows of data consisting of 700 good and 300 bad, with 21 variables (V1 to V20) while V21 is the target or output. To analyze the data, 10 model scenarios were built to identify the best model. The model results show that the model performance is getting better after SMOTE is carried out with the accuracy and AUC values increasing. The best model was obtained in scenario 1 (90% train data and 10 test data) with an accuracy value of 0.83 and AUC 0.918. The model evaluation results show that the XGBoost algorithm can be used to analyze credit data before accepting/rejecting a credit application.

Credit, XGBoost, Knowledge Discovery in Database, Data Mining

Kata Kunci:

(*) Korespondensi:

1910631170063@student.unsika.ac.id

How to Cite: Saputra, A., Sari, B., & Rozikin, C. (2024). Penerapan Algoritma Extreme Gradient Boosting (Xgboost) Untuk Analisis Risiko Kredit. *Jurnal Ilmiah Wahana Pendidikan*, 10(7), 27-36. <https://doi.org/10.5281/zenodo.10960080>

PENDAHULUAN

Perekonomian di negara ini ditopang oleh banyak sektor, salah satunya adalah sektor keuangan yang memiliki peranan yang sangat penting. Misalnya membantu pemerintah untuk membiayai proyek-proyek infrastruktur negara. Selain itu, sektor keuangan juga berperan penting dalam menggerakkan perekonomian masyarakat dengan menawarkan produk jasa keuangan mereka berupa produk kredit kendaraan bermotor, kredit usaha, KUR, dan masih banyak lagi. Berdasarkan UU No. 10 Tahun 1998 (UU Perbankan), kredit adalah penyediaan uang atau persamaan dengan itu, berdasarkan perjanjian antara kreditur dan debitur yang mewajibkan pihak peminjam untuk melunasi utangnya beserta bunga setelah jangka waktu tertentu. Otoritas Jasa Keuangan (OJK) telah mengeluarkan data outstanding pinjam online dari bulan Maret 2022 sampai Maret 2023.

Berdasarkan data *outstanding* pinjam *online* tersebut, pada bulan Maret tahun ini kredit tidak lancar (30-90 hari) melonjak sebesar Rp3,58 triliun (71,13% YoY) dengan pinjaman perseorangan tertinggi sebesar Rp3,3 triliun dan selebihnya

dari badan usaha. Kemudian yang mengalami kredit macet melebihi 90 hari pada bulan Maret 2023 sebesar Rp1,43 triliun dari total sisa pinjaman yang belum terbayar oleh debitur. Nilai ini meningkat jauh sebesar Rp1,35 triliun (65,33%) dibandingkan pada periode sebelumnya (*YoY*), di mana sebesar Rp1,14 triliun adalah kredit macet dari perseorangan dan sisanya adalah badan usaha. Hal ini terjadi karena berawal dari kurangnya keakuratan dalam proses *screening customer* (Usriyati et al., 2022).

Kredit macet adalah kredit dengan perjanjian kontrak yang tidak dapat diselesaikan oleh debitur. Hal ini terjadi ketika debitur mengalami kesulitan keuangan, gagal membayar angsuran atau gagal memenuhi kewajiban pembayaran lainnya dalam jangka waktu yang telah ditentukan (Donald R. van Deventer, Kenji Imai, 2013). Kredit macet sering disebabkan oleh analisis data yang tidak teliti dalam proses pinjaman (tidak akurat dalam mengevaluasi data nasabah/ pelanggan, yang menyebabkan kesalahan dalam pemberian pinjaman). Dengan berkembangnya ilmu pengetahuan, banyak perbankan kini telah mengadopsi teknologi untuk meningkatkan akurasi analisis keputusan kredit, termasuk *data mining*.

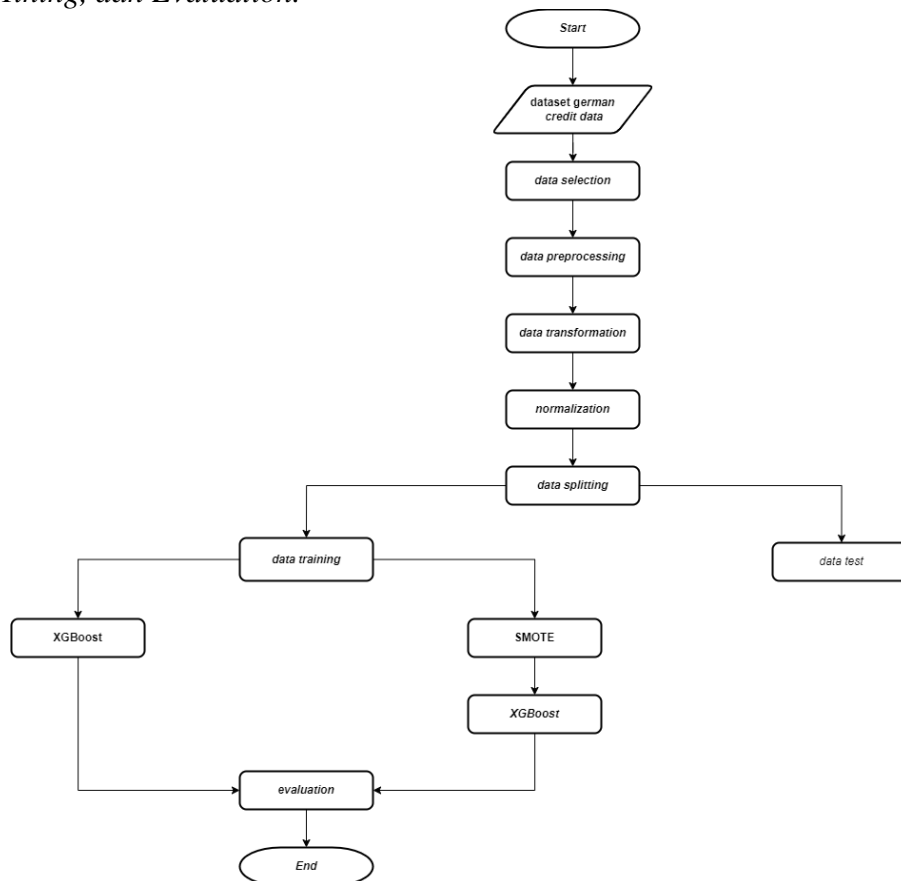
Berdasarkan penelitian sebelumnya yang dilakukan oleh (Prasojo & Haryatmi, 2021) tentang prediksi kelayakan pemberian kredit menggunakan algoritma *random forest*, didapatkan hasil klasifikasi pada model prediksi dengan akurasi 83% sehingga termasuk pada nilai *very good model* (80-89%). Penelitian berikutnya oleh (Givari et al., 2022) mengenai analisis persetujuan kredit dengan membandingkan algoritma SVM, *random forest* dan XGBoost. Hasil penelitian ini menyimpulkan model XGBoost memperoleh nilai tertinggi pada nilai *accuracy* 82%, *recall* 70%, dan *precision* 92%. Penelitian lainnya oleh (Delima et al., 2023) dengan masalah yang sama tetapi menggunakan algoritma XGBoost dan Adaptive Boosting, menghasilkan algoritma XGBoost lebih baik untuk model klasifikasi pemberian pinjaman dengan nilai AUC 0,92 dan algoritma AdaBoost adalah 0,89.

XGBoost (*Extreme Gradient Boosting*) adalah pengembangan lebih lanjut dari algoritma *Gradient Tree Boosting* berbasis *ensemble*, yang dapat secara efektif menangani masalah *machine learning* dengan skala yang lebih besar (Herni Yulianti et al., 2022). XGBoost dikembangkan dengan pengoptimalan 10 kali lebih cepat daripada metode peningkatan gradien lainnya (Chen & Guestrin, 2016). XGBoost memberikan keunggulan dalam hal kinerja, efisiensi waktu, dan memori yang telah digunakan di beberapa bidang penelitian dimulai dari bidang kesehatan, risiko kredit, dan metagenomik (Wang et al., 2020).

Penelitian ini menerapkan algoritma XGBoost dalam menganalisis risiko kredit dalam upaya untuk mengoptimasi proses *screening customer*. Penelitian ini bertujuan memberikan informasi untuk membantu pemberi pinjaman dalam menyusun strategi bisnis khususnya dalam bidang penyaluran kredit kepada debitur secara tepat sehingga dapat mengurangi risiko-risiko kredit di waktu yang akan datang.

METODE

Metode atau langkah-langkah untuk mencapai tujuan dari penelitian ini adalah metodologi *Knowledge Discovery in Database (KDD)*, yang terdiri dari lima langkah yaitu *Data Selection, Data Preprocessing, Data Transformation, Data Mining, dan Evaluation*.



Gambar 1 Flowchart Rancangan Penelitian Metodologi KDD

Tahap pertama yaitu menyeleksi dan mengambil data di *UCI Machine Learning Repository* yaitu *german credit data*. Ditahap *data selection* ini dilakukan deskripsi data seperti melihat jumlah data dan tipe data. Tahap kedua adalah *data preprocessing* yang dilakukan dengan eksplorasi data dan pembersihan data seperti menangani data yang tidak sesuai, menangani data yang hilang, dan menangani *outlier*. Tahap selanjutnya adalah *data transformation* yaitu proses pengkodean data untuk mengubah skala data asli ke bentuk lain. Pada fase ini dilakukan normalisasi data dengan metode *RobustScaler* dan menyeimbangkan data menggunakan metode SMOTE.

Langkah selanjutnya adalah penambahan data menggunakan algoritma XGBoost untuk mendapatkan pola atau informasi data yang sudah dipilih. Proses *data mining*. Adapun 5 skenario yang akan dilakukan untuk mendapatkan hasil terbaik dapat dilihat pada Tabel 1.

Tabel 1 Skenario Pembagian *Dataset*

No.	Skenario	Algoritma XGBoost	Algoritma XGBoost + SMOTE
	A (90:10)	Model 1	Model 1

	B (80:20)	Model 2	Model 2
	C (70:30)	Model 3	Model 3
	D (60:40)	Model 4	Model 4
	E (50:50)	Model 5	Model 5

Tahap terakhir pada metodologi KDD adalah evaluasi. Pada tahap ini dilakukan perubahan pola menjadi informasi yang mudah dipahami. Tahapan ini untuk mengetahui pola atau informasi yang diperoleh apakah sesuai dengan tujuan sebelumnya untuk mengevaluasi performa algoritma XGBoost dengan menggunakan *confusion matrix*, dan melihat nilai AUC dari kurva ROC.

HASIL & PEMBAHASAN

Hasil

Data Selection

Dalam kumpulan data ini, setiap orang diklasifikasikan menjadi risiko kredit baik atau buruk menurut serangkaian atributnya. Pada tahap ini dilakukan proses deskripsi data. Deskripsi data yang dilakukan terdiri dari penjelasan jumlah data dan tipe data. *Dataset* berjumlah 1000 *row* data yang terdiri dari 700 *good* (1) dan 300 *bad* (2). Sebagian data dapat dilihat pada Tabel 2 dan tipe data pada Gambar 2.

Tabel 2 *Dataset German Credit*

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	...	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21
0	A11	6	A34	A43	1169	A65	A75	4	A93	A101	...	A121	67	A143	A152	2	A173	1	A192	A201	1
1	A12	48	A32	A43	5951	A61	A73	2	A92	A101	...	A121	22	A143	A152	1	A173	1	A191	A201	2
2	A14	12	A34	A46	2096	A61	A74	2	A93	A101	...	A121	49	A143	A152	1	A172	2	A191	A201	1
3	A11	42	A32	A42	7882	A61	A74	2	A93	A103	...	A122	45	A143	A153	1	A173	2	A191	A201	1
4	A11	24	A33	A40	4870	A61	A73	3	A93	A101	...	A124	53	A143	A153	2	A173	2	A191	A201	2

Pada tabel diatas terlihat bahwa data terdiri dari 21 variable mulai (V1 sampai dengan V20) sedangkan V21 merupakan target atau *output*.

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 21 columns):
#   Column      Non-Null Count  Dtype
---  -
0    V1          1000 non-null   object
1    V2          1000 non-null   int64
2    V3          1000 non-null   object
3    V4          1000 non-null   object
4    V5          1000 non-null   int64
5    V6          1000 non-null   object
6    V7          1000 non-null   object
7    V8          1000 non-null   int64
8    V9          1000 non-null   object
9    V10         1000 non-null   object
10   V11         1000 non-null   int64
11   V12         1000 non-null   object
12   V13         1000 non-null   int64
13   V14         1000 non-null   object
14   V15         1000 non-null   object
15   V16         1000 non-null   int64
16   V17         1000 non-null   object
17   V18         1000 non-null   int64
18   V19         1000 non-null   object
19   V20         1000 non-null   object
20   V21         1000 non-null   int64
dtypes: int64(8), object(13)
memory usage: 164.2+ KB
```

Gambar 2 Tipe *Dataset*

Pada Gambar 2 terlihat tipe data *object* ada 13 kolom dan 8 kolom bertipe *integer*.

Data Preprocessing

Tahapan selanjutnya adalah data preprocessing, yang melibatkan beberapa langkah penting dalam menangani kualitas data sebelum proses data mining. Berikut adalah urutan tahapan yang diperlukan:

1. Pengecekan tipe *Dataset*

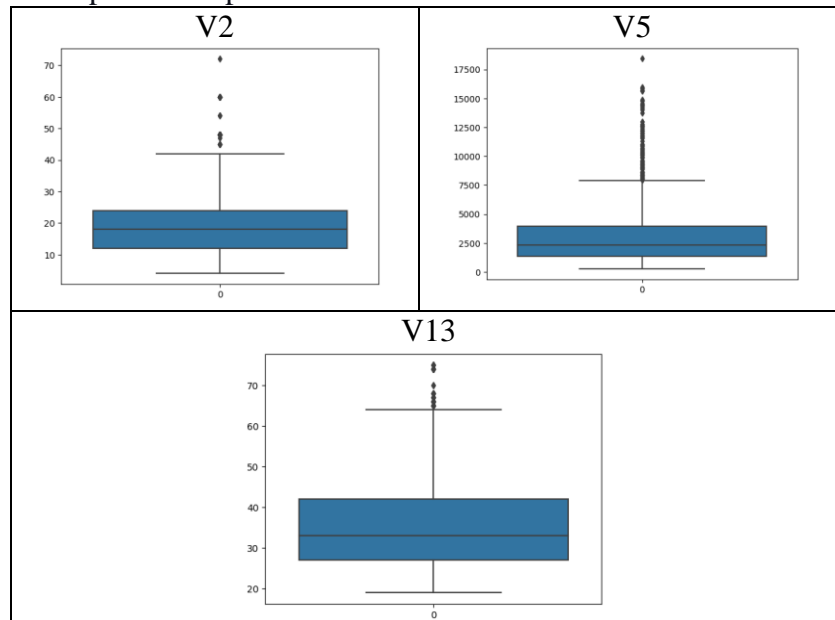
Memastikan tipe *dataset* yang benar sangat penting dilakukan karena metode analisis dan pemrosesan data yang akan digunakan dapat berbeda tergantung pada tipe data yang ada. Tipe data yang benar juga dapat mempengaruhi bagaimana *outlier* dideteksi atau diatasi dalam *dataset*. Misalnya, pada data numerik, *outlier* dapat diidentifikasi berdasarkan nilai-nilai yang berada di luar rentang tertentu. Setelah dilakukan pengecekan, tidak ada ketidaksesuaian tipe *dataset* sehingga data dapat terhindar dari kesalahan dalam menganalisis data.

2. Pengecekan *Missing Value*

Sumber dari *dataset* pada penelitian ini sudah menjelaskan tidak ada missing value dan pada tahap ini tidak perlu dilakukan *handling missing value*.

3. Pengecekan *Outlier*

Selanjutnya dalam tahap data *preprocessing* yaitu melakukan pengecekan *outlier*. Langkah ini dilakukan untuk mengecek apakah ada nilai yang terlalu besar atau kecil di luar batas normal yang diharapkan, seperti nilai yang jauh lebih tinggi atau rendah dari nilai rata-rata atau median pada kolom numerik. Hasil pengecekan dapat dilihat pada Gambar 3.



Gambar 3 Hasil Pengecekan *Outlier*

Berdasarkan Gambar 3 didapati hasil dari pengecekan data *outlier* menggunakan fungsi *boxplot()* dengan bantuan dari *library Pandas* dan *Matplotlib*, terdapat beberapa kolom yang memiliki nilai ekstrem yaitu pada kolom “V2”, “V5”, dan “V13” yang dianggap sebagai *outlier*. Berdasarkan hasil tahap data *preprocessing* ini, *outlier* yang ditampilkan bukan hasil kesalahan pengukuran atau entri data yang salah, sehingga tetap dipertahankan untuk tahap selanjutnya. Contoh hasil pengecekan pada variabel V2 (durasi pinjaman) terdapat 1 data dengan pengajuan durasi pinjaman selama 72 bulan, pada variabel V5 (jumlah kredit) jumlah kredit

diatas 15000 hanya ada 5 data yang nilainya juga berbeda, kemudian pada variabel V13(umur) terdapat hanya 2 yang berumur 75 tahun.

Data Transformation

Proses transformasi data bertujuan untuk memudahkan proses pemodelan dengan data yang digunakan. Jika dilihat pada Gambar 2 tertera tipe data dari kolom yang berupa kategorikal, sedangkan pemodelan hanya bisa jika data bertipe numerik, maka perlu di transformasi lagi agar data diubah menjadi numerik. Berdasarkan sumber *dataset* yang diperoleh juga menunjukkan data belum seimbang untuk dilakukan pemodelan yakni 700 data berlabel *good* dan 300 data berlabel *bad*. Tahap transformasi ini juga dilakukan *handling imbalance data*. Langkah pertama yang dilakukan data yang bertipe kategori harus diubah terlebih dahulu. Perubahan tipe *dataset* dapat dilihat pada Tabel 3.

Tabel 3 Dataset Setelah Transformasi

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	...	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21
0	1	6	5	4	1169	5	5	4	3	1	...	1	67	3	2	2	3	1	2	1	1
1	2	48	3	4	5951	1	3	2	2	1	...	1	22	3	2	1	3	1	1	1	0
2	4	12	5	7	2096	1	4	2	3	1	...	1	49	3	2	1	2	2	1	1	1
3	1	42	3	3	7882	1	4	2	3	3	...	2	45	3	3	1	3	2	1	1	1
4	1	24	4	1	4870	1	3	3	3	1	...	4	53	3	3	2	3	2	1	1	0

Setelah dilakukan proses transformasi terhadap dataset didapatkan hasil sesuai Tabel 3. Proses transformasi selanjutnya normalisasi data dengan menggunakan metode *RobustScaler*. Normalisasi adalah proses mengubah skala data sehingga semua variabel berada dalam rentang tertentu. Berdasarkan pada Tabel 2 dan Gambar 2, terlihat bahwa atribut yang bertipe kategorikal mengalami perubahan yang dapat dilihat di Tabel 3 dan Tabel 4.

Tabel 4 Dataset Setelah Normalisasi Menggunakan RobustScaler

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
680	0.666667	-1.00	0.0	2.0	-0.273883	0.0	-1.0	-1.0	-1.0	0.0	-0.5	0.5	1.533333	0.0	0.0	0.0	0.0	0.0	0.0	0.0
177	-0.333333	-1.00	1.0	0.5	-0.686254	1.0	2.0	0.5	0.0	0.0	0.5	0.0	1.266667	0.0	0.0	1.0	0.0	0.0	0.0	0.0
395	0.000000	1.75	0.5	2.0	3.238832	0.5	1.0	-0.5	0.0	0.0	0.0	0.5	-0.066667	0.0	-1.0	0.0	0.0	0.0	1.0	0.0
911	0.000000	0.50	1.0	0.0	0.825086	0.0	-1.0	-0.5	-1.0	0.0	0.5	0.0	-0.533333	-2.0	0.0	0.0	-1.0	0.0	0.0	0.0
793	0.333333	0.50	0.0	0.0	0.191409	0.0	2.0	0.0	-2.0	0.0	0.5	0.5	1.200000	0.0	1.0	0.0	0.0	0.0	0.0	0.0

RobustScaler pada dasarnya adalah metode normalisasi yang rentang nilai dari data tetap terjaga dengan mengubah skala data. *RobustScaler* menggabungkan median dan *interquartile* untuk perhitungannya sehingga nilai *outlier* tidak terlalu berpengaruh. Pemusatan dan penskalaan terjadi secara independen pada setiap fitur dengan menghitung statistik yang relevan pada sampel di set pelatihan.

Langkah berikutnya *handling imbalance data* dengan metode SMOTE. *Synthetic Minority Over-sampling* (SMOTE) adalah metode yang digunakan dalam pembelajaran mesin untuk mengatasi masalah ketidakseimbangan kelas dalam kumpulan data. Ketidakseimbangan kelas terjadi ketika satu kelas (biasanya kelas minoritas) kurang terwakili secara signifikan dibandingkan dengan kelas lain dalam suatu masalah klasifikasi. Data yang sebelumnya tidak

seimbang setelah dilakukan *handling imbalance data* menggunakan metode SMOTE data menjadi sama antara label yang bernilai 1 dengan yang bernilai 0. Di mana label yang bernilai 0 oleh metode SMOTE dibuat sampel sintesis sebanyak 400. Sehingga data yang sebelumnya 300 ditambah 400 data sintesis menjadi 700 data (seimbang dengan kelas mayoritas).

Data Mining

Pada tahap *data mining*, desain pengujian menggunakan *dataset* yang telah dipersiapkan sebelumnya. *Dataset* yang digunakan memuat 1000 data entri dengan total 21 kolom. *Dataset* tersebut telah dipartisi menjadi set data *training* dan data *testing* dengan perbandingan tertentu, di mana dalam skenario 1 data di *split* menjadi 5 skenario dengan masing-masing menerapkan *handling imbalance data* menggunakan metode SMOTE, dan skenario 2 *split* data tanpa menerapkan *handling imbalance data* dengan 5 skenario yang sama. Dalam skenario A menggunakan perbandingan 90% *data training* dan 10% *data testing*, skenario B menggunakan perbandingan 80% *data training* dan 20% *data testing*, skenario C menggunakan perbandingan 70% *data training* dan 30% *data testing*, skenario D menggunakan perbandingan 60% *data training* dan 40% *data testing*, skenario E menggunakan perbandingan 50% *data training* dan 50% *data testing*.

Dengan menentukan 5 skenario dapat menghemat waktu komputasi, semakin banyak skenario yang diuji maka semakin lama waktu yang dibutuhkan untuk melatih dan menguji model. Pembagian *dataset* ini dilakukan dengan menggunakan fungsi *train_test_split* dari *library sklearn*. Dalam langkah ini, penggunaan fungsi *train_test_split* untuk membagi *dataset* menjadi set pelatihan dan pengujian dengan perbandingan tertentu merupakan bagian dari strategi analisis yang dipilih.

Evaluation

Pada tahap evaluasi ini, dilakukan penilaian terhadap semua model yang telah dibangun untuk mengidentifikasi model yang paling cocok atau optimal. Evaluasi dilakukan dengan menggunakan metrik atau kriteria tertentu yang relevan. Proses evaluasi bertujuan untuk mengevaluasi performa dan kualitas model dalam memprediksi atau mengklasifikasikan data baru. Dengan membandingkan hasil dari berbagai model yang dibangun, dapat diidentifikasi model mana yang memiliki performa yang lebih baik dan sesuai dengan tujuan atau kebutuhan penelitian. Penilaian model dilakukan dengan membandingkan hasil prediksi model dengan nilai sebenarnya dari data uji atau data validasi. Performa model dibandingkan dalam bentuk tabel yang dapat dilihat pada Tabel 5 dan Tabel 6.

Tabel 5 Akurasi Tanpa Metode SMOTE Dan Dengan Metode SMOTE

Skenario	Tanpa SMOTE	Dengan SMOTE
1	0,74	0,83
2	0,75	0,83
3	0,77	0,79
4	0,74	0,79
5	0,74	0,76

Tabel 5 menunjukkan akurasi model dari setiap kondisi dan skenario. Dapat diketahui bahwa model yang dibangun tanpa menerapkan metode SMOTE menunjukkan akurasi yang rendah. Jika dibandingkan dengan menerapkan metode SMOTE hasilnya menunjukkan peningkatan nilai akurasi.

Tabel 6 Nilai AUC Tanpa Metode SMOTE Dan Dengan Metode SMOTE

Skenario	Tanpa SMOTE	Dengan SMOTE
1	0,822	0,918
2	0,792	0,905
3	0,773	0,887
4	0,775	0,872
5	0,764	0,857

Tabel 6 menunjukkan nilai AUC model dari setiap kondisi dan skenario, AUC digunakan untuk mengukur performansi model dalam memisahkan kelas. Model yang dibangun tanpa menerapkan metode SMOTE menunjukkan nilai AUC yang rendah dibandingkan dengan menerapkan metode SMOTE

Pembahasan

Semakin tinggi nilai AUC menunjukkan performansi model juga semakin bagus dalam memisahkan kelas. Berikut rentang nilai AUC dalam pengkategorian performansi model pada Tabel 7.

Tabel 7 Kategori Model Berdasarkan Rentang Nilai AUC

No	Rentang Nilai AUC	Kategori
1	0,9 - 1,0	<i>Very Good</i>
2	0,8 - 0,9	<i>Good</i>
3	0,7 - 0,8	<i>Fair</i>
4	0,6 - 0,7	<i>Poor</i>
5	0,5 - 0,6	<i>Fail</i>

Hasil evaluasi model terbaik berdasarkan kurva ROC, diperoleh nilai AUC model XGBoost dengan penambahan metode SMOTE adalah skenario 1 dan 2 yang termasuk dalam kategori *very good* dalam memisahkan kedua kelas dengan tepat. Artinya model yang dihasilkan dari skenario 1 dan 2 dapat digunakan untuk menguji kelayakan pemberian kredit. Di ikuti nilai AUC yang diperoleh dari skenario 3, 4, dan 5 yang termasuk kategori *good* dalam memisahkan kedua kelas dengan tepat. Ini berarti skenario 3, 4, dan 5 layak digunakan untuk menguji kelayakan pemberian kredit.

KESIMPULAN

Berdasarkan analisis yang dilakukan menggunakan algoritma XGBoost, menunjukkan bahwa algoritma XGBoost sangat baik dalam menganalisis data kredit. Dalam tahap *data mining*, 10 model dibangun untuk mengidentifikasi model yang terbaik. Sebelum melakukan pemodelan pada *german credit data*, penyeimbangan kelas data dilakukan terlebih dahulu dengan SMOTE.

Hasil evaluasi menunjukkan bahwa penggunaan algoritma XGBoost dapat digunakan untuk menganalisis data kredit sebelum menerima/ menolak pengajuan kredit. Hasil evaluasi model juga menunjukkan bahwa performa model semakin baik setelah dilakukan SMOTE dengan nilai *accuracy* dan nilai AUC yang mengalami peningkatan. Model terbaik didapatkan pada skenario 1 (*data train 90%* dan *10 data test*) dengan nilai *accuracy* 0,83 dan AUC 0,918.

REFERENCES

- Bandyopadhyay, A. (2016). Managing portfolio credit risk in banks. *Managing Portfolio Credit Risk in Banks*, 1–361. <https://doi.org/10.1017/CBO9781316550915>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-Aug*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Cherif, I. L., & Kortebi, A. (2019). On using eXtreme Gradient Boosting (XGBoost) Machine Learning algorithm for Home Network Traffic Classification. *IFIP Wireless Days, 2019-April*, 1–6. <https://doi.org/10.1109/WD.2019.8734193>
- Delima, R., Hosianna, M., Pebrianty, D., Amalia, J., & Analysis, A. C. R. (2023). *Credit Risk Analysis dengan Algoritma Extreme Gradient Boosting dan Adaptive Boosting. 05*, 1–7.
- Donald R. van Deventer, Kenji Imai, and M. M. (2013). *Advanced Financial Risk Management: Tools and Techniques for Integrated Credit Risk and Interest Rate Risk Management*.
- Elmayati, E. (2017). Data Mining Dengan Metode Clustering Untuk Pengolahan Informasi Persediaan Obat Pada Klinik Srikandi Medika Berbasis Web. *Pelita Informatika: Informasi Dan Informatika*, 16(4), 357–362. <https://ejurnal.stmik-budidarma.ac.id/index.php/pelita/article/view/531/482>
- Enry, U., & Umaidah, Y. (2022). *Data Analytics* (p. Mandala).
- Ginting, I. F. P., Saripurna, D., & Fitriani, E. (2021). Penerapan Data Mining Dalam Menentukan Pola Ketersediaan Stok Barang Berdasarkan Permintaan Konsumen Di Chykes Minimarket Menggunakan Algoritma Apriori. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 20(1), 28. <https://doi.org/10.53513/jis.v20i1.2504>
- Givari, M. R., Sulaeman, M. R., & Umaidah, Y. (2022). Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit. *Nuansa Informatika*, 16(1), 141–149. <https://doi.org/10.25134/nuansa.v16i1.540>

- Ha, J., Kambe, M., & Pe, J. (2011). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- Hanif, I. (2020). *Implementing Extreme Gradient Boosting (XGBoost) Classifier to Improve Customer Churn Prediction*. <https://doi.org/10.4108/eai.2-8-2019.2290338>
- Herni Yulianti, S. E., Oni Soesanto, & Yuana Sukmawaty. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *Journal of Mathematics Theory and Application*, 4(1), 21–26. <https://doi.org/10.31605/jomta.v4i1.1792>
- Kadiyala, A., & Kumar, A. (2018). Applications of Python to evaluate environmental data science problems. In *Environmental Progress and Sustainable Energy* (Vol. 36, Issue 6, pp. 1580–1586). John Wiley and Sons Inc. <https://doi.org/10.1002/ep.12786>
- Marceau, L., Qiu, L., Vandewiele, N., & Charton, E. (2019). *A comparison of Deep Learning performances with other machine learning algorithms on credit scoring unbalanced data*. *ML*. <http://arxiv.org/abs/1907.12363>
- Miranda, A., & Muslim Lhaksmana, K. (2022). Classification Analysis of Waiting Period for Telkom University Alumni to Get Jobs Using Decision Tree and Support Vector Machine. *Technology and Science (BITS)*, 4(2). <https://doi.org/10.47065/bits.v4i2.1963>
- Prasojo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(2), 79–89. <https://doi.org/10.25077/teknosi.v7i2.2021.79-89>
- Pushner, D. C., & Bouteille, S. (2013). *The Handbook of Credit Risk Management: Originating, Assessing, and Managing Credit Exposures*.
- Putri, D. Y., Andreswari, R., & Hasibuan, M. A. (2022). *Analysis of Students Graduation Target Based on Academic Data Record Using C4.5 Algorithm Case Study: Information Systems Students of Telkom University*.
- Usriyati, R., Priyono, N., & Khabibah, N. A. (2022). Penanganan Kredit Bermasalah (Non Performing Loan) Pada PT. BPR BKK Muntilan (Perseroda). *Akuntansi Dan Manajemen*, 17(1), 60–71. <https://doi.org/10.30630/jam.v17i1.176>
- Virantika, E., Kusnawi, K., & Ipmawati, J. (2022). Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia. *Jurnal Media Informatika Budidarma*, 6(3), 1657. <https://doi.org/10.30865/mib.v6i3.4325>
- Wang, C., Deng, C., & Wang, S. (2020). Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. *Pattern Recognition Letters*, 136, 190–197. <https://doi.org/10.1016/j.patrec.2020.05.035>