ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

MODEL PREDIKSI RISIKO KREDIT MACET MENGGUNAKAN XGBOOST DENGAN AI EXPLAINER BERBASIS SHAP

Alvian Maulana Mansyur✉, Putu Manik Prihatini, Gde Brahupadhya Subiksa
Jurusan Teknologi Informasi, Politeknik Negeri Bali, Badung, Indonesia
Email: alvianmaulana0704@gmail.com

DOI: https://doi.org/10.46880/jmika.Vol10No2.pp635-644

ABSTRACT

Credit is a financial instrument that carries the risk of default when borrowers are unable to fulfill their repayment obligations. This risk may lead to financial losses for lending institutions, making accurate credit risk prediction essential. This study aims to develop a credit risk prediction model using Extreme Gradient Boosting (XGBoost) and point out its predictions using an AI explainer based on SHapley Additive exPlanations (SHAP). The dataset was obtained from the internal database of a fintech company in Indonesia and consisted of 116,250 records with 19 attributes. The research stages included Exploratory Data Analysis (EDA), missing value and outlier handling, data splitting, model training, evaluation, and result interpretation. The XGBoost model achieved an accuracy of 92%. For the Less Current class, the model obtained a precision of 0.91, a recall of 0.97, and an F1-score of 0.94. For the current class, the model achieved a precision of 0.86, a recall of 0.62, and an F1-score of 0.72. For the Default class, the model obtained a precision of 0.97, a recall of 0.91, and an F1-score of 0.94. SHAP analysis as an AI explainer revealed that the income and side_income features were the most influential factors in determining credit risk.

Keyword: Credit Risk, Classification, XGBoost, SHAP.

ABSTRAK

Kredit merupakan salah satu instrumen keuangan yang memiliki risiko gagal bayar ketika debitur tidak mampu memenuhi kewajiban pembayarannya. Risiko ini dapat menyebabkan kerugian finansial bagi lembaga pemberi pinjaman, sehingga prediksi risiko kredit yang akurat menjadi sangat penting. Penelitian ini bertujuan untuk membangun model prediksi risiko kredit menggunakan Extreme Gradient Boosting (XGBoost) serta menjelaskan hasil prediksinya menggunakan AI Explainer berbasis SHapley Additive exPlanations (SHAP). Dataset yang digunakan diperoleh dari basis data internal sebuah perusahaan fintech di Indonesia dan terdiri atas 116.250 data dengan 19 atribut. Tahapan penelitian meliputi Exploratory Data Analysis (EDA), penanganan missing value dan outlier, pembagian data, pelatihan model, evaluasi, serta interpretasi hasil. Model XGBoost menghasilkan akurasi sebesar 92%. Pada kelas Kurang Lancar, model memperoleh nilai precision sebesar 0,91, recall sebesar 0,97, dan F1-score sebesar 0,94. Pada kelas Lancar, model memperoleh nilai precision sebesar 0,86, recall sebesar 0,62, dan F1-score sebesar 0,72. Pada kelas Macet, model memperoleh nilai precision sebesar 0,97, recall sebesar 0,91, dan F1-score sebesar 0,94. Analisis SHAP sebagai AI Explainer menunjukkan bahwa fitur income dan side_income merupakan faktor yang paling berpengaruh dalam menentukan risiko kredit.

Kata Kunci: Risiko Kredit, Klasifikasi, XGBoost, SHAP.

PENDAHULUAN

Kredit adalah salah satu instrumen keuangan yang dapat digunakan untuk mendorong kemajuan perekonomian suatu negara baik dari perbankan maupun non-perbankan atau perusahaan yang bergerak di bidang keuangan digital (*fintech*). Hambatan yang menjadi tantangan utama yang harus dihadapi adalah adanya risiko kredit macet atau *non-performing loans* (NPL). Kredit macet adalah kondisi di saat debitur atau peminjam tidak dapat mengembalikan pembayaran yang seharusnya dibayarkan yang diberikan oleh bank maupun penyedia layanan kredit lainnya(Yanda & Rahmazaniati, 2024). Faktor terjadinya kredit macet sangat bervariasi dan kompleks, mulai dari faktor internal seperti *overcrediting* sehingga peminjam tidak bisa membayar kembali kelebihan utang yang diajukan. Tidak hanya itu, kesalahan dalam menganalisis juga menjadi faktor pendorong terjadinya kredit macet. Faktor eksternal dapat terjadi ketika peminjam mendapatkan musibah yang tak terduga, baik bencana langsung dari alam maupun non-alam (Rafaella & Prabowo, 2021; Yanda & Rahmazaniati, 2024; Zakiati & Sinaga, 2025).

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Perkembangan teknologi di Indonesia terutama di bidang *financial and technology* atau *fintech* cukup memberikan banyak perubahan terutama pada bagian sistem yang awalnya bersifat manual kini banyak yang sudah beralih ke otomatisasi yang mengedepankan efisiensi baik dari segi pengeluaran hingga jangkauan akses yang bisa diberikan kepada masyarakat yang membuat pertumbuhan ekonomi di Indonesia menjadi lebih inklusif (Jange et al., 2024). Salah satu upaya yang telah dilakukan di banyak perusahaan *fintech* terutama lembaga pemberi kredit di Indonesia adalah klasifikasi kredit yang dibangun untuk meminimalisir terjadinya kredit macet dan dapat digunakan untuk menganalisis kesehatan pembayaran nasabah.

Layanan klasifikasi kredit pada nasabah yang dilakukan oleh lembaga pemberi kredit terutama di Indonesia masih mengandalkan metode tradisional seperti pembobotan statis terhadap karakteristik data pribadi dan demografi pengguna. Hal ini menyebabkan banyaknya data pengguna yang harus digunakan untuk menentukan klasifikasi kredit seseorang, sehingga waktu yang dibutuhkan untuk mendapatkan keluaran tentang klasifikasi kredit cenderung kurang efisien. Oleh karena itu Penelitian ini dilakukan untuk menguji efektivitas prediksi risiko kredit berbasis pembelajaran mesin (*machine learning*) menggunakan data nasabah yang diperoleh secara sekunder pada suatu perusahaan tanpa melibatkan data tambahan dari penyedia data pihak ketiga.

Beberapa penelitian terdahulu telah dilakukan dalam upaya mengidentifikasi terjadinya kredit macet berbasis *machine learning* seperti prediksi risiko kredit yang telah dilakukan dengan membandingkan SVM, *Random Forest* dan XGBoost dengan dataset yang diperoleh melalui Kaggle (Givari et al., 2022). Hasil dari penelitian tersebut disimpulkan bahwa XGBoost menghasilkan nilai metrik evaluasi yang cukup unggul dibandingkan SVM dan Random Forest. Penelitian lain juga melakukan pengujian untuk membandingkan model Random Forest dan XGBoost untuk memprediksi risiko kredit menggunakan dua jenis dataset yang berjumlah 10.000 dan 100.000 data (Dachi & Sitompul, 2023). Hasil yang dapat disimpulkan adalah XGBoost memberikan hasil dengan akurasi yang tinggi dibandingkan Random Forest.

Beberapa penelitian yang sudah dilakukan menunjukkan performa yang baik pada model XGBoost. Namun terdapat celah penelitian (*research gap*) dimana sebagian penelitian yang telah dilakukan masih bergantung pada pada data sekunder *open source* dari Kaggle yang kurang merepresentasikan perilaku debitur serta pola ketidakseimbangan data yang sering terjadi terutama lembaga pemberi kredit di Indonesia. Kemudian terdapat masalah baru dimana penelitian terdahulu masih bersifat *black box* yang artinya kurangnya transparansi yang membongkar bagaimana model dapat menyimpulkan suatu keputusan risiko.

Mengatasi permasalahan tersebut, penelitian ini mengusulkan pendekatan yang berorientasi pada performa dan transparansi model. Penelitian ini menerapkan algoritma XGBoost karena terbukti dari penelitian terdahulu bahwa model ini mampu mengambil keputusan kredit yang akurat. Penerapan algoritma XGBoost juga akan memanfaatkan penggabungan metode *Stratified K-Fold Cross Validation* untuk mengoptimasi model dan teknik *oversampling* menggunakan SMOTE untuk memitigasi terjadinya *imbalance data* pada dataset. Selain itu, SHAP dipilih sebagai *explainable AI* (XAI) yang menjelaskan karakteristik serta keterkaitan dari setiap fitur secara transparan yang dapat digunakan untuk meningkatkan rasa kepercayaan terhadap keputusan hasil prediksi risiko kredit dari model yang dibangun.

TINJAUAN PUSTAKA

Risiko Kredit

Risiko kredit adalah kondisi seseorang yang melakukan peminjaman namun sang peminjam tidak dapat mengembalikan kewajibannya untuk membayar kembali pinjaman yang diajukan yang telah disepakati pada perjanjian awal. Risiko kredit yang buruk memiliki dampak yang buruk juga bagi kesehatan peminjaman nasabah sehingga kredit yang disalurkan harus lebih diperhatikan (Assa & Loindong, 2023).

XGBoost

Extreme Gradient Boosting atau XGBoost merupakan algoritma *machine learning* yang bisa digunakan untuk klasifikasi dan regresi. XGBoost bekerja menggunakan gabungan dari banyak *weak learner* seperti *Decision Tree* yang bertujuan untuk meminimalisir kesalahan prediksi melalui penyempurnaan setiap iterasi pohon keputusan. XGBoost merupakan model turunan yang menyempurnakan metode *Gradient Boosting* yang menambahkan regularisasi yang digunakan untuk mencegah terjadinya *overfitting* (Aydin & Ozturk, 2021; Qinghe et al., 2022).

SMOTE

SMOTE atau *Synthetic Minority Oversampling Technique* adalah metode yang sering digunakan untuk menyeimbangkan kelas pada data yang distribusinya tidak seimbang dengan tujuan agar model dapat mengumpulkan informasi lebih banyak mengenai kelas fitur yang lebih sedikit jumlahnya. SMOTE

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

memanfaatkan teknik *oversampling* yang ditentukan berdasarkan *k-nearest neighbor* untuk menghasilkan data sintetik yang memiliki karakteristik yang hampir mirip dengan data sebenarnya (Hermaliani & Ernawati, 2024).

SKCV

SKCV atau singkatan dari *Stratified K-Fold Cross Validation* merupakan turunan dari metode K-Fold. SKCV bekerja sangat baik terutama pada dataset yang sifatnya tidak seimbang. SKCV bekerja dengan membagi dataset kedalam jumlah *folds* yang ditentukan, namun mempertahan proporsi dari dataset sehingga seluruh karakteristik data akan diperlakukan secara adil (Qiu, 2024).

AI Explainer

AI Explainer atau *Explainable AI* (XAI) adalah sebuah konsep yang bertujuan untuk meningkatkan transparansi untuk mengatasi permasalahan yang membantu pengguna dalam memahami keluaran model yang menghasilkan akurasi tinggi namun sulit untuk dipahami, sehingga keluaran model yang kompleks dapat disajikan dalam bentuk yang lebih mudah untuk dipahami (Hoffman et al., 2023).

SHAP

SHAP atau *Shapley Additive Exlaination* adalah sebuah metode yang digunakan untuk menjelaskan dan mengukur kontribusi masing-masing fitur *machine learning* yang digunakan dengan cara mengkalkulasi keterkaitan dan kontribusi setiap fitur yang digunakan untuk prediksi. SHAP menggunakan teori permainan kooperatif. Namun bukan permainan secara harfiah melainkan menggunakan organisasi selayaknya permainan, dimana fitur akan diperlakukan sebagai pemain. Pemain akan dinilai kontribusinya terhadap suatu model yang akan diperlakukan sebagai skor (Huang et al., 2024; Reddy, 2024).

Metrik Evaluasi

Metrik evaluasi adalah metode yang digunakan untuk mengukur kemampuan model dalam melakukan prediksi. Umumnya jenis metrik evaluasi yang sering digunakan dalam menilai kinerja model adalah *recall, accuracy, precision*, *F1-Score* dan *confusion matrix.*

Recall atau sering disebut sebagai *true positive rate* adalah penilaian yang digunakan untuk mengetahui seberapa baik model dapat memahami contoh yang relevan (Allam et al., 2025. *Recall* ditulis dalam persamaan (1).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negatives} \times 100 \quad (1)$$

Keterangan:
Recall = Nilai *recall* yang akan dicari.
True Positive = Jumlah data yang diprediksi positif dan relevan positif.
False Negative = Jumlah data yang diprediksi negatif dan relevan negatif.

Accuracy adalah penilaian terhadap rasio keberhasilan model memprediksi suatu contoh yang belum pernah dilihat sebelumnya (Allam et al., 2025. *Accuracy* ditulis dalam persamaan (2).

$$Accuracy = \frac{True\ Positive + True\ Negatives}{Total\ Samples} \quad (2)$$

Keterangan:
Accuracy = Nilai accuracy yang akan dicari.
True Positive = Jumlah data yang diprediksi positif dan relevan positif.
False Negative = Jumlah data yang diprediksi negatif tapi sebenarnya positif.
Total Samples = Jumlah total seluruh data di dataset.

Precision adalah penilaian untuk mengukur akurasi prediksi yang bernilai positif. *Precision* juga sering disebut sebagai nilai prediktif positif (Allam et al., 2025) *Precision* ditulis dalam persamaan (3).

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \times 100 \quad (3)$$

Keterangan:
Precision = Nilai *precision* yang akan dicari.
True Positive = Jumlah data yang diprediksi positif dan relevan positif.
False Positive = Jumlah data yang diprediksi positif tapi sebenarnya negatif (tidak relevan).

F1-Score adalah penilaian yang menilai keharmonisan hasil rata-rata antara *precision* dan *recall*. *F1-Score* sangat berguna dalam skenario dengan distribusi kelas yang tidak merata (Allam et al., 2025). *F1-Score* dihitung dalam persamaan (4).

$$F1 - Score = 2\left(\frac{Precision \times Recall}{Precision + Recall}\right) \quad (4)$$

Keterangan:
F1-Score = Nilai *F1-Score* yang akan dicari.
Precision = Nilai *precision* yang telah ditentukan.
Recall = Nilai *Recall* yang telah ditentukan.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Confusion Matrix adalah matriks N×N, dimana N mewakili baris jumlah *output* dalam bentuk numerik. Tujuannya adalah menyederhanakan hasil evaluasi agar lebih mudah dipahami, seperti menjelaskan berapa jumlah data yang berhasil diprediksi dan berapa jumlah data yang gagal diprediksi. *Confusion matrix* sering digunakan dalam klasifikasi untuk menilai performa model (Swaminathan & Tantri, 2024).

Penelitian Sebelumnya

Penelitian yang telah dilakukan untuk melakukan prediksi risiko kredit dengan membandingkan kinerja model *Logistic Regression*, *Random Forest*, *XGBoost* dan *AdaBoost* (Lin, 2024). Pengujian ini menggunakan dataset yang berasal dari Kaggle. Hasil dari pengujian ini mengatakan XGBoost mampu memberikan akurasi tertinggi dibanding model lainnya dengan perolehan akurasi model mencapai 93.26% di ikuti *precision* 0.93 dan *recall 0.98*, sehingga *F1-Score* yang dihasilkan mencapai 0.81. Sedangkan model dengan akurasi terendah dihasilkan oleh *Logistic Regression* dengan akurasi mencapai 81.20% diikuti *precision 0.81* dan *recall 0.98*.

Penelitian lain juga telah dilakukan dengan membandingan kinerja model *Support Vector Machine* (SVM), *Random Forest* dan XGBoost untuk memprediksi kelayakan kredit seseorang (Givari et al., 2022). Dataset yang digunakan berasal dari Kaggle dan bersifat *open source*. Pengujian ini mengukur keunggulan model menggunakan nilai *precision*, *recall* dan *accuracy*. Hasil dari penelitian ini adalah SVM dan *Random Forest* mampu memberikan hasil yang serupa, dimana nilai *precision*, *recall* dan *accuracy* menghasilkan nilai yang sama yaitu 0.50. Sedangkan pada model XGBoost memberikan hasil yang cukup signifikan membaik dengan nilai *precision* sebesar 0.92, *recall* sebesar 0.70 dan *accuracy* sebesar 0.82.

Penelitian lain juga dilakukan dengan melakukan perbandingan model XGBoost dan *Random Forest* dengan dua dataset yang berjumlah 10.000 data dan 100.000 data (Dachi & Sitompul, 2023). Hasil yang dicapai adalah *Random Forest* mampu memberikan nilai *F1-Score* 0.70 pada data berjumlah 10.000 dan *F1-Score* 0.995 pada dataset berjumlah 100.000. Model XGBoost mampu memberikan nilai *F1-Score* yang sangat sempurna, yaitu dengan nilai 1.0 pada dataset yang berjumlah 10.000 dan 100.000 data.

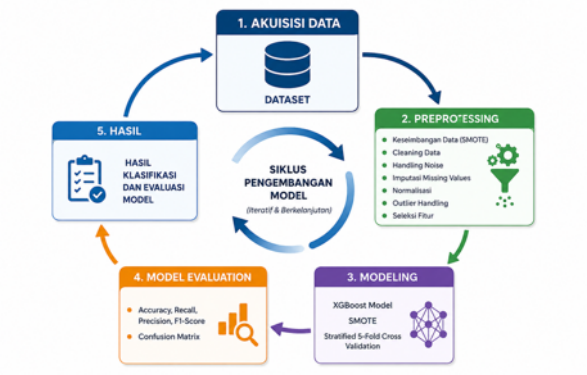
Penelitian lain dilakukan untuk mengklasifikasi kelayakan pinjaman menggunakan perbandingan model XGBoost dan *Logistic Regression* (Yaqin & Ramadhani, 2023). Data yang digunakan berasal dari penilaian kredit Australia yang jumlah dan fiturnya tidak disebutkan. Hasil pengujian ini XGBoost mampu menghasilkan akurasi sebesar 85.51% dengan *precision* 83.80%, *recall* 84.04% dan *F1-Score* 83.81%. Sedangkan *Logistic Regression* memiliki performa yang lebih baik dengan akurasi sebesar 85.94% diikuti *precision* 80.08%, *recall* 91.52% dan *F1-Score* 85.35% yang menandakan *Logistic Regression* menjadi yang terbaik dari pebandingan 2 metode ini dan XGBoost hanya unggul dari segi *precision*.

Berdasarkan pemaparan penelitan terdahulu, model XGBoost secara konsisten menunjukkan performa yang sangat baik dalam memprediksi risiko kredit. Namun terdapat beberapa celah penelitian (*research gap*) dimana mayoritas penelitian yang sudah dilakukan sangat bergantung pada dataset sekunder bersifat *open sources* dari Kaggle yang karakteristiknya telah melalui proses kurasi yang ketat. Kondisi ini cenderung gagal dalam mencerminkan perilaku debitur serta ketidakseimbangan kelas (*class imbalance*) ekstrem yang sering ditemui pada industri finansial terutama di Indonesia. Masalah lain yang dapat ditemukan adalah aspek interpretabilitas model. Meskipun XGBoost terbukti mendominasi dan menghasilkan akurasi yang tinggi, penelitian yang dilakukan masih bersifat *black box* atau kurangnya transparansi bagaimana model dapat mengambil keputusan tertentu yang dapat menurunkan rasa kepercayaan pengguna dalam memitigasi risiko kredit.

METODE PENELITIAN

Jenis Penelitian

Penelitian ini dilakukan dengan jenis pendekatan terapan (*applied research*) yang memiliki alur pemecahan masalah yang terstruktur dan sistematis. Alur pemecahan masalah yang akan dilakukan meliputi akuisisi data, *preprocessing data*, *splitting dataset*, transformasi data, *train model*, evaluasi model dan *explanation AI* menggunakan SHAP. Tahapan penelitian ini dapat dilihat pada Gambar 1 di bawah ini.



Gambar 1. Tahap Penelitian

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Akuisisi Data

Akuisisi data dilakukan menggunakan dataset sekunder yang diperoleh dari salah satu perusahaan *fintech* di Indonesia yang identitasnya tidak diungkap untuk menjaga kerahasiaan institusi dan data. Dataset yang digunakan memiliki jumlah 116.250 baris data dengan 19 atribut meliputi; *data_id*, *age*, *gender, marital_status*, *occupation*, *plafon*, *loan_type*, *purpose*, *tenor*, *tenor_type*, *interest_rate, income*, *side_income*, *collateral_value*, *max_days_past_due*, *total_days_past_due*, *DBR*, *LTV* dan *worst_collectability*.

Preprocessing Data

Preprocessing dataset penting untuk dilakukan dan merupakan hal krusial yang dapat mempengaruhi kinerja model, sehingga dataset perlu disesuaikan dengan kebutuhan model XGBoost melalui tahapan sebagai berikut:

1. *Data Cleaning*: Tahap ini dilakukan dengan menghapus kolom fitur duplikat. Kolom fitur yang kosong dilakukan teknik imputasi dengan mengisi data *null* dengan *mean* untuk numerik dan kata "*unknown*" untuk fitur kategorikal.
2. *Data Splitting*: Dataset dibagi dengan perbandingan 80:20 yang berarti 80% atau sebanyak 93.494 baris data akan digunakan sebagai data latih sedangkan 20% atau 22.874 baris sisanya merupakan data pengujian.
3. *Outlier Handling*: Setelah dataset terbagi, fitur numerik seperti *age, tenor* dan *interest_rate* pada data latih dibersihkan dari *outlier* atau nilai ekstrem seperti nilai yang sangat besar atau angka negatif yang tidak masuk akal. Nilai yang dianggap *outlier* ini diatasi menggunakan imputasi yang di isi dengan *mean* dari total fitur.
4. Transormasi Fitur: Fitur dataset yang bertipe *string* seperti fitur *gender*, *marital_status*, *occupation*, *loan_type*, *purpose* dan *tenor_type* ditransformasi menggunakan *ordinal encoder* menjadi numerik agar dataset ini bisa diproses dalam perhitungan bobot oleh model XGBoost.

Model *Pipeline*

Proses pelatihan model diawali dengan membuat *pipeline* yang akan berisi model XGBoost sebagai model utama. Model XGBoost dibangun menggunakan pengujian *hyperparameter* pada Tabel 1 berikut.

Tabel 1. Hyperparameter XGBoost

No	***Hyperparameter***	Nilai Uji
1	*n_estimators*	1000
2	*max_depth*	12
3	*learning_rate*	0.01
4	*min_child_weight*	5
5	*reg_alpha*	0
6	*reg_lambda*	0.5

Berdasarkan *hyperparameter* pengujian yang ditampilkan pada Tabel 1, nilai *n_estimators* yang lebih besar dibanding nilai *learning_rate* sangat mempengaruhi kinerja model (Ørebaek & Geitle, 2021), sehingga ditetapkan 1000 sebagai nilai *n_estimators* dan 0.01 sebagai nilai *learning_rate*. Selanjutnya, ditetapkan nilai 12 untuk *max_depth* dan nilai 5 sebagai pengaturan *min_child_weight*. Untuk regularisasi model, ditetapkan nilai 0 sebagai *reg_alpha* serta nilai 0.5 sebagai nilai *reg_lambda* untuk mencegah *overfitting*.

SMOTE yang digunakan untuk mengatasi data *imbalance* dengan menentukan nilai *K-Nearest Neighbor* dengan nilai 10 atau K=10 dengan alasan nilai K=10 merupakan nilai yang stabil dalam SKCV (Nti et al., 2021). Kemudian model dioptimasi dengan *Stratified K-Fold Cross Validation* (SKCV) dengan nilai *fold* adalah 5 atau *K-Fold*=5 yang menandakan model akan dilatih sebanyak 5 kali.

Model *Evaluation*

Penilaian kinerja XGBoost dinilai berdasarkan model evaluasi *recall, accuracy, precision* dan *F1-Score*. Semakin tinggi nilai yang dihasilkan, maka semakin baik kinerja model. Penilaian kinerja XGBoost juga didukung menggunakan *confusion matrix* untuk memberikan gambaran jelas bagaimana model mampu menebak benar prediksi dan mengetahui jumlah kesalahan prediksi yang dilakukan model XGBoost.

SHAP *AI Explainer*

SHAP digunakan untuk menjelaskan fitur apa yang paling mempengaruhi XGBoost dalam mengidentifikasi risiko kredit macet menggunakan visualisasi *beeswarm* dengan pengurutan secara vertikal untuk fitur terbaik.

HASIL DAN PEMBAHASAN

Exploratory Data Analysis

Hasil dari *Exploratory Data Analysis* atau EDA ditemukan masih terdapat banyak *missing value* dan *outlier* sehingga sebelum melakukan pelatihan model, dataset harus melalui *preprocessing* agar dataset menjadi lebih optimal ketika digunakan oleh model XGBoost. Pada Tabel 2 disajikan jumlah data di setiap fitur yang memiliki *missing value* dan *outlier*.

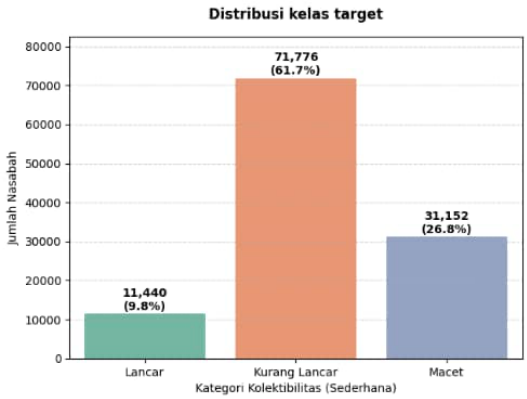
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Tabel 2. Perincian Missing Value dan Outlier Pada Setiap Fitur

Fitur	*Missing Value* (%)	*Outlier* (%)
age	1.3	4.46
plafon	0.0	9.51
purpose	33.97	0.0
tenor	0.0	2.7
interest_rate	0.0	5.91
side_income	0.0	0.23
collateral_value	61.75	0.0
DBR	0.0	9.71
LTV	0.0	38.25

Selain temuan *missing value* dan *outlier* yang ditampilkan pada Tabel 2, hasil EDA juga menunjukkan distribusi data pada fitur target *worst_collectability* pada dataset memiliki karakteristik yang *imbalance* yang ditampilkan pada grafik distribusi data pada Gambar 2 berikut ini.



Gambar 2. Distribusi Data Imbalance Pada Target

Berdasarkan grafik yang ditunjukkan pada Gambar 2. Nasabah yang tergolong lancar berjumlah 11.440 atau 9.8% dari seluruh dataset. Nasabah yang tergolong kurang lancar berjumlah 71.776 atau 61.7% dari seluruh dataset dan nasabah yang tergolong macet berjumlah 31.152 atau 26.8% dari seluruh dataset. Hasil ini menunjukkan dataset bersifat *imbalance* atau data memiliki ketidakseimbangan distribusi data

Preprocessing Data

Berdasarkan hasil *preprocessing* yang berhasil dilakukan, dataset telah dibersihkan dan fitur kategorikal yang telah ditransformasi sesuai dengan kebutuhan model XGBoost disajikan melalui 10 data teratas fitur kategorikal pada Tabel 3.

Tabel 3. Data Teratas Fitur Kategorikal Setelah Transformasi

gender	*marital status*	*loan type*	*purpose*	*tenor type*	*occupation*
0.00	3.00	51.00	68.00	0.00	11.00
1.00	3.00	26.00	110.00	0.00	2.00
0.00	3.00	26.00	110.00	0.00	2.00
0.00	3.00	51.00	110.00	0.00	12.00
1.00	3.00	32.00	89.00	0.00	11.00
0.00	3.00	13.00	51.00	0.00	8.00
1.00	0.00	26.00	110.00	0.00	1.00
1.00	3.00	32.00	61.00	0.00	2.00
1.00	3.00	28.00	53.00	0.00	12.00
0.00	3.00	28.00	61.00	0.00	2.00

Berdasarkan Tabel 3, fitur kategorikal berhasil diubah dari data bertipe *string* menjadi numerik yang menandakan dataset ini sudah mengikuti algoritma XGBoost dan siap digunakan sebagai data latih.

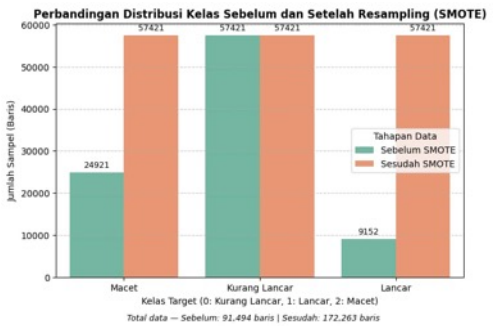
Data yang sudah bersih kemudian dibagi menjadi data latih dan data uji dengan perbandingan 80:20, yang menghasilkan 91.494 data latih dan 22.874 data uji. *Data split* disajikan pada Tabel 4 lengkap dengan distribusi data pada 3 kelas target.

Tabel 4. Data Split

	K_L (%)	M (%)	L (%)	Total data
Data latih	62.8	27.2	1	91.494
Data uji	62.8	27.2	1	22.878

Keterangan:
K_L = Kurang Lancar
M = Macet
L = Lancar

Tabel 4 menunjukkan terdapat 62.8% data yang terklasifikasi kurang lancar, 27.2% data macet dan 1% data lancar. Hal ini merupakan data *imbalance* yang ekstrem. Untuk itu, perlu dilakukan resampling data menggunakan teknik SMOTE untuk meningkatkan kelas minoritas menjadi setara dengan kelas mayoritas.



Gambar 3. Distribusi Data setelah Resampling (SMOTE)

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

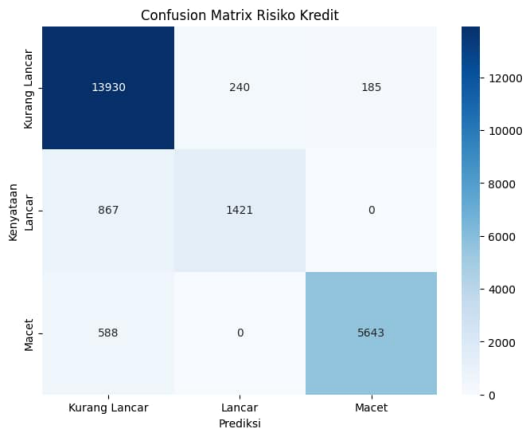
ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Gambar 3 menunjukkan terjadinya perubahan signifikan distribusi data setelah diterapkan teknik SMOTE. Batang hijau merepresentasikan data sebelum menggunakan teknik SMOTE dan batang oranye merepresentasikan kelas setelah menggunakan SMOTE. Hasilnya kelas (0) atau kelas Macet berjumlah 24.921 data dan meningkat menjadi 57.421 data setelah dilakukan *resampling*. Kelas (1) yang mewakili kelas Kurang Lancar merupakan kelas mayoritas berjumlah 57.421 data dan kelas ini tidak di *resampling*. Kelas (2) atau kelas Lancar yang awalnya berjumlah 9.152 berhasil ditingkatkan secara sintesis hingga masing-masing menyamai volume kelas mayoritas pada angka 57.421 data, sehingga membentuk total keseluruhan 172.263 baris data latih.

Pembentukan rasio 1:1:1 ini mengubah cara model XGBoost mempelajari data dengan memberikan bobot penalti yang adil pada setiap iterasinya dan membuat model tidak terjebak dalam akurasi yang paradoks.

Model Evaluation

Confusion matrix pada Gambar 4 menunjukkan bahwa model mampu memprediksi kelas mayoritas yaitu Kurang Lancar sebanyak 13.930 data yang bernilai *true positive* atau benar dan total 425 data yang salah prediksi atau *false negative*. Model juga mampu membedakan karakteristik nasabah pada data minoritas seperti lancar dan macet. Hasilnya model mampu memprediksi 1.421 data untuk *true positive* dengan *false negative* di angka 867 data. Model juga mampu membedakan nasabah yang macet dengan *true positive* sebanyak 5.643 data namun memiliki *false negative* sebanyak 588 data.



Gambar 4. Confusion Matrix

Confusion matrix pada Gambar 4 juga diperkuat dengan nilai *precision*, *recall, accuracy* dan *F1-Score* yang cukup baik. Kelas Kurang Lancar (0) model mampu menghasilkan *precision* sebesar 0.91, *recall* sebesar 0.97 dan *F1-Score* sebesar 0.94. Hasil ini menandakan model mampu memprediksi kelas mayoritas dengan sangat efektif dengan tingkat *false negative* yang sangat rendah.

Kelas Lancar (1) yang menjadi kelas minoritas ekstrem juga bisa diprediksi oleh model XGBoost yang menghasilkan *precision* sebesar 0.86, *recall* 0.62 dan *F1-Score* sebesar 0.72. Fenomena rendahnya nilai *recall* ini dipicu oleh status awal kelas Lancar sebagai kelas minoritas ekstrem yang memiliki jumlah data paling sedikit sebelum proses *resampling*. Eksploitasi sintesis data yang begitu besar oleh SMOTE menyebabkan efek tumpang tindih (*overlapping*) spasial antar fitur terhadap kelas lainnya. Hal ini mengakibatkan model cenderung menghasilkan prediksi yang bias, sehingga ketika dihadapkan data yang asli model cenderung bersikap konservatif dan melewatkan beberapa sampel nasabah lancar yang sebenarnya (true positive)

Kelas Macet (2) mampu diprediksi dengan baik oleh model. Hal ini dapat dibuktikan dari nilai *precision* sebesar 0.97, *recall* sebesar 0.91 dan *F1-Score* sebesar 0.94 yang menandakan bahwa model juga sangat efektif dalam memprediksi nasabah dengan prediksi macet. Secara keseluruhan model mampu melakukan prediksi dengan akurasi mencapai 0.92, namun masih terdapat sedikit kesalahan selama proses pelatihan model. Performa model XGBoost secara keseluruhan ditampilkan pada Tabel 5.

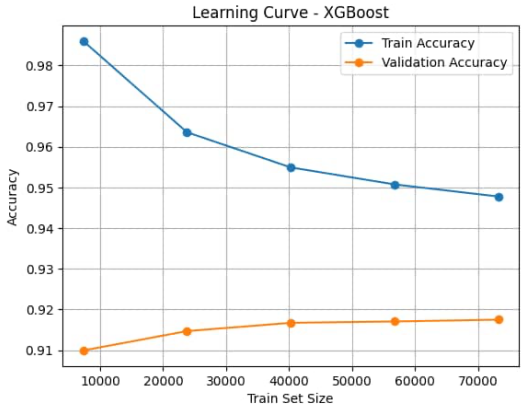
Tabel 5. Classification Report

	precision	***recall***	***f1-score***	***support***
Kurang Lancar	0.91	0.97	0.94	14355
Lancar	0.86	0.62	0.72	2288
Macet	0.97	0.91	0.94	6231
accuracy			0.92	

Berdasarkan performa model yang disajikan pada Tabel 5 terbukti bahwa SMOTE mampu meningkatkan dan mengatasi permasalahan data *imbalance* pada dataset. Selain itu, proses pelatihan model menggunakan SKCV juga mampu memberikan kinerja yang cukup baik, dibuktikan dari *accuracy* model secara keseluruhan yang mencapai nilai 0.92 atau 92%. Hasil ini juga diperkuat dengan hasil *learning curve* yang ditampilkan pada Gambar 5.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

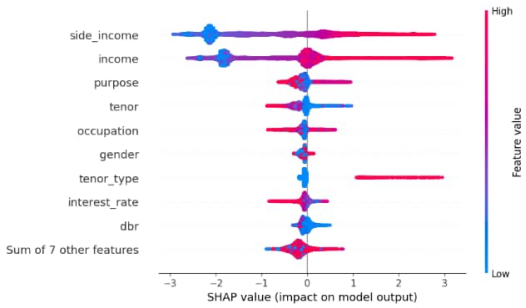


Gambar 5. Learning Curve – XGBoost

Berdasarkan *learning curve* pada Gambar 5 menunjukkan visualisasi pergerakan *learning curve* selama pelatihan model. Terjadi penurunan pada *train accuracy* pada saat pelatihan karena bertambahnya jumlah data yang dipelajari. Namun *validation accuracy* cenderung menaik dan stabil di akhir. Hal ini menunjukkan grafik *learning curve* yang bergerak secara konvergen, meskipun terdapat sedikit gap sebanyak 0.03 atau 3% di akhir pelatihan, dimana nilai akhir *train accuracy* berada di angka 0.94 dan nilai akhir *validation accuracy* di angka 0.91. Namun hal ini secara empiris menandakan model memiliki kemampuan generalisasi yang baik dan tidak menunjukkan gejala *overfitting* yang signifikan.

SHAP *AI Explainer*

Transparansi yang menjelaskan bagaimana model dapat menentukan prediksi berdasarkan fitur apa yang paling mempengaruhi model dalam mengklasifikasikan nasabah ditunjukkan menggunakan *beeswarm* pada Gambar 6. Sumbu *y* (vertikal) merupakan fitur yang diurutkan dan sumbu *x* (horizontal) menunjukkan dampak terhadap model.



Gambar 6. SHAP Beeswarm

Dari Gambar 6 dapat disimpulkan bahwa fitur *side income* dan *income* memiliki pengaruh dominan terhadap pengambilan keputusan model. Pada fitur *side income* dan *income,* terlihat sebaran data berwarna merah bergerak memanjang ke arah positif hingga angka 3 yang menandakan nasabah dengan fitur pendapatan dan pendapatan sampingan yang tinggi mampu menaikan probabilitas risiko kredit lancar. Sebaliknya fitur *tenor* dan *interest_rate* justru memiliki hubungan negatif terhadap prediksi kelas lancar, yang mengidentifikasi bahwa tenor yang panjang dan bunga yang tinggi justru menurunkan probabilitas kelas lancar. Fitur lain justru tersebar hanya di dekat angka 0 yang menandakan kecilnya pengaruh yang diberikan terhadap prediksi tersebut. Hal ini menunjukkan bahwa model sangat bergantung pada demografis pendapatan atau keadaan finansial nasabah dalam melakukan prediksi risiko kredit.

KESIMPULAN

Penelitian ini terbukti berhasil membangun model prediksi risiko kredit menggunakan XGBoost dengan baik dan mampu menangani *data imbalance* menggunakan gabungan metode SMOTE dan SKCV. Secara keseluruhan, model XGBoost mampu menghasilkan performa prediktif yang optimal dengan nilai akurasi total mencapai 92%. Validitas performa ini didukung oleh grafik learning curve yang bergerak konvergen dengan jarak (gap) yang sangat sempit, yang membuktikan bahwa model memiliki kemampuan generalisasi yang tinggi serta terbebas dari masalah overfitting maupun underfitting Meskipun model mengalami kesulitan dalam mengidentifikasi nasabah yang memiliki risiko lancar dengan nilai *precision* 0.86, *recall* 0.62 dan *F1-Score* 0.72. Rendahnya nilai *recall* ini dipicu oleh status awal kelas Lancar sebagai kelas minoritas ekstrem. manipulasi sintesis data yang terlampau masif oleh SMOTE menyebabkan *overlapping* pada batasan keputusan (*decision boundary*) model. Walaupun demikian, model sangat efektif dalam mengidentifikasi nasabah kurang lancar dan macet. Model secara signifikan menghasilkan *precision* 0.91, *recall* 0.97 dan *F1-Score* 0.94 untuk identifikasi nasabah kurang lancar. Model juga menghasilkan nilai *precision* 0.97, *recall* 0.91 dan *F1-Score* 0.94 untuk identifikasi nasabah yang macet. Implementasi SHAP juga terbukti efektif dalam mengidentifikasi fitur *income* dan *side_income* sebagai fitur yang paling mempengaruhi pengambilan keputusan model sekaligus menyelesaikan permasalahan sifat *black box* oleh model XGBoost.

Meskipun XGBoost dinilai cukup baik dalam melakukan prediksi risiko kredit, penanganan kelas *imbalance* cenderung masih belum maksimal, walaupun sudah menggabungkan metode SMOTE dan SKCV. Oleh karena itu, penelitian berikutnya disarankan untuk berfokus pada eksplorasi terkait

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

penanganan ketidakseimbangan dataset yang ekstrem. Penerapan teknik *K-Fold* dan SMOTEENN direkomendasikan untuk menangani kasus ketidakseimbangan data yang ekstrem dan eksplorasi lebih dalam terkait *hyperparameter* dan pengujian model lainnya seperti *Balanced Random Forest* dan *Cost-Sensitive Logistic Regression* yang berfokus pada pembobotan kelas minoritas.

DAFTAR PUSTAKA

Allam, H., Makubvure, L., Gyamfi, B., Graham, K. N., & Akinwolere, K. (2025). Text Classification: How Machine Learning Is Revolutionizing Text Categorization. *Information*, *16*(2), 130. https://doi.org/10.3390/info16020130

Assa, V., & Loindong, S. S. R. (2023). Analisis Pengaruh Risiko Kredit, Kecukupan Modal Dan Likuiditas Terhadap Kinerja Keuangan Pada Bank Bumn Di Bursa Efek Indonesia (BEI). *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi*, *11*(4), 1048–1057. https://doi.org/10.35794/emba.v11i4.51747

Aydin, Z. E., & Ozturk, Z. K. (2021). Performance analysis of XGBoost classifier with missing data. *Manchester Journal of Artificial Intelligence and Applied Sciences (MJAIAS), 2*(02), 2021.

Dachi, J. M. A. S., & Sitompul, P. (2023). Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit. *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam*, *2*(2), 87–103. https://doi.org/10.55606/jurrimipa.v2i2.1470

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.v16i1.5406

Ridwan, R., Hermaliani, E. H., & Ernawati, M. (2024). Penerapan: Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian. *Computer Science (Co-Science), 4*(1), 80-88. https://doi.org/10.31294/coscience.v4i1.2990

Hoffman, R. R., Mueller, S. T., Klein, G., Jalaeian, M., & Tate, C. (2023). Explainable AI: roles and stakeholders, desirements and challenges. *Frontiers in Computer Science*, *5*, 1117848. https://doi.org/10.3389/fcomp.2023.1117848

Huang, Y., Zhou, Y., Chen, J., & Wu, D. (2024). Applying Machine Learning and SHAP Method to Identify Key Influences on Middle-School Students' Mathematics Literacy Performance. *Journal of Intelligence*, *12*(10), 93. https://doi.org/10.3390/jintelligence12100093

Jange, B., Pendi, I., & Susilowati, E. M. (2024). Peran Teknologi Finansial (Fintech) dalam Transformasi Layanan Keuangan di Indonesia. *Indonesian Research Journal on Education*, *4*(3), 1199–1205. https://doi.org/10.31004/irje.v4i3.1007

Lin, J. (2024, January). Research on loan default prediction based on logistic regression, randomforest, xgboost and adaboost. In *SHS web of conferences* (Vol. 181, p. 02008). EDP Sciences. https://doi.org/10.1051/shsconf/202418102008

Nti, I. K., Nyarko-Boateng, O., & Aning, J. (2021). Performance of Machine Learning Algorithms with Different K Values in K-fold CrossValidation. *International Journal of Information Technology and Computer Science*, *13*(6), 61–71. https://doi.org/10.5815/ijitcs.2021.06.05

Ørebæk, O. E., & Geitle, M. (2021, March). Exploring the Hyperparameters of XGBoost Through 3D Visualizations. In *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*. http://ceur-ws.org

Qinghe, Z., Wen, X., Boyan, H., Jong, W., & Junlong, F. (2022). Optimised extreme gradient boosting model for short term electric load demand forecasting of regional grid system. *Scientific Reports*, *12*(1), 19282. https://doi.org/10.1038/s41598-022-22024-3

Qiu, J. (2024). An Analysis of Model Evaluation with Cross-Validation: Techniques, Applications, and Recent Advances. *Advances in Economics, Management and Political Sciences*, *99*(1), 69–72. https://doi.org/10.54254/2754-1169/99/2024ox0213

Rafaella, A. C., & Prabowo, B. (2021). Analisis Kredit Macet pada Kredit Usaha Rakyat (KUR) PT. Bank Rakyat Indonesia (Persero) Tbk Unit Rungkut Surabaya pada Masa Pandemi Covid-19. *Al-Kharaj: Jurnal Ekonomi, Keuangan & Bisnis Syariah*, *4*(2), 368–379. https://doi.org/10.47467/alkharaj.v4i2.674

Reddy, A. K. (2024). Bridging AI and Human Understanding: Interpretable Deep Learning in Practice. *Journal of Informatics Education and Research*, *4*(3). https://doi.org/10.52783/jier.v4i3.2200

Swaminathan, S., & Tantri, B. R. (2024). Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, *27*, 4023–4031. https://doi.org/10.53555/AJBR.v27i4S.4345

Yanda, N., & Rahmazaniati, L. (2024). Analisis Faktor-Faktor yang Mempengaruhi Kredit Macet Pada Bank BTN Kantor Cabang Pembantu Syariah Meulaboh. *Innovative: Journal Of Social Science Research*, *4*(2), 545–557. https://doi.org/10.31004/innovative.v4i2.9488

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Yaqin, A. (2023). Penilaian kredit menggunakan algoritma XGBoost dan logistic regression. *Jurnal Informatika: Jurnal Pengembangan IT, 8*(1), 4-10.

Zakiati, A., & Sinaga, O. S. (2025). Faktor yang Menyebabkan Kredit Macet pada Debitur di Bank BRI Unit Pasar Baru Balikpapan. *Jurnal Hukum, Politik dan Ilmu Sosial*, *4*(2), 54–65. https://doi.org/10.55606/jhpis.v3i3.3937

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.