

# Prediksi Risiko Kredit Nasabah Menggunakan Algoritma Data Mining: Studi Kasus pada PT Toyota Astra Finance

DOI: <http://dx.doi.org/10.35889/progresif.v21i2.2909>

Creative Commons License 4.0 (CC BY – NC)



Icha Winadya Permadani<sup>1</sup>, Raka Sulisty<sup>2</sup>, Muhammad Fadli<sup>3\*</sup>, Erliyan Redy Susant<sup>4</sup>

<sup>1,2,3,4</sup>Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

<sup>3</sup>Ekonomi dan Bisnis, Politeknik Negeri Lampung, Bandar Lampung, Indonesia

\*e-mail *Corresponding Author*: [muhammadfadlofficial@polinela.ac.id](mailto:muhammadfadlofficial@polinela.ac.id)

## Abstract

*This study aims to develop a credit risk prediction model for customers at PT Toyota Astra Financial Services using data mining algorithms, specifically Random Forest and XGBoost. In response to the challenge of non-performing loans (NPL), machine learning-based predictive models offer an effective solution to identify potential risks early. The research utilizes historical customer data encompassing demographic information, employment status, and loan history. After data preprocessing, the models were evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. The results indicate that XGBoost outperformed other models with an accuracy of 91.67% and an F1-score of 0.89 for the positive class. These findings demonstrate that applying machine learning algorithms can significantly enhance credit selection efficiency and reduce potential losses from defaulted loans.*

**Keywords:** Credit Risk; Machine learning; Random Forest; XGBoost; Data mining.

## Abstrak

Penelitian ini bertujuan untuk membangun model prediksi risiko kredit nasabah pada PT Toyota Astra Financial Services dengan memanfaatkan algoritma *data mining*, khususnya *Random Forest* dan *XGBoost*. Dalam menghadapi tantangan kredit macet, model prediktif berbasis *machine learning* dapat memberikan solusi yang efektif untuk mengidentifikasi potensi risiko sejak dini. Penelitian ini menggunakan data historis nasabah yang mencakup informasi demografi, status pekerjaan, dan riwayat pinjaman. Setelah melalui tahap pra-pemrosesan data, model dievaluasi menggunakan metrik akurasi, *presisi*, *recall*, *F1-score*, dan *ROC-AUC*. Hasil menunjukkan bahwa *XGBoost* memiliki performa terbaik dengan akurasi sebesar 91,67% dan *F1-score* 0,89 pada kelas positif. Temuan ini menunjukkan bahwa penerapan algoritma *machine learning* dapat meningkatkan efisiensi seleksi kredit dan mengurangi potensi kerugian akibat kredit bermasalah.

**Kata kunci:** Risiko Kredit; Machine learning; Random Forest; XGBoost; Data mining

## 1. Pendahuluan

Industri pembiayaan kendaraan bermotor di Indonesia mengalami pertumbuhan signifikan seiring meningkatnya kebutuhan masyarakat terhadap kendaraan roda dua maupun roda empat. Pertumbuhan ini berdampak pada meningkatnya jumlah penyaluran kredit oleh perusahaan pembiayaan. Namun, peningkatan penyaluran kredit juga diikuti oleh munculnya risiko kredit bermasalah atau *Non Performing Loan* (NPL) yang dapat merugikan perusahaan dan menurunkan kepercayaan pelanggan maupun investor. Oleh karena itu, topik mengenai prediksi risiko kredit menjadi penting untuk diteliti karena berhubungan langsung dengan keberlangsungan dan stabilitas industri pembiayaan. [1] [2].

Salah satu perusahaan pembiayaan terkemuka di sektor ini adalah PT Toyota Astra Finance (TAF). Sebagai bagian dari ekosistem layanan keuangan Toyota, TAF memberikan berbagai fasilitas kredit kepada nasabah, mulai dari pembiayaan awal, *top-up*, hingga *refinancing*.

Seiring dengan pertumbuhan jumlah nasabah, tantangan yang dihadapi adalah meningkatnya potensi kredit bermasalah. Berdasarkan data internal industri, NPL menjadi indikator utama kualitas portofolio kredit, dan jika tidak dikelola dengan baik dapat menimbulkan kerugian finansial yang signifikan. Masalah ini bersifat terukur karena NPL memiliki ambang batas toleransi yang ditetapkan oleh regulator, sehingga perusahaan harus mampu menjaga kualitas kredit agar tetap sehat.

Penilaian risiko kredit secara tradisional masih banyak bergantung pada pengalaman analis, yang sering kali subjektif dan kurang efektif dalam menangani data berukuran besar dan kompleks. Seiring perkembangan teknologi informasi, pendekatan berbasis *data mining* dan *machine learning* dinilai lebih objektif, cepat, dan akurat dalam memprediksi risiko kredit. Berbagai algoritma seperti Regresi Logistik, *Decision Tree*, dan *Random Forest* telah banyak digunakan dalam penelitian sebelumnya untuk memprediksi kelayakan kredit, dengan hasil yang menunjukkan peningkatan akurasi prediksi dibanding metode konvensional [3][4][5]. *Random Forest*, misalnya, terbukti mampu mengatasi masalah *overfitting* dan bekerja baik pada data dengan variabel numerik maupun kategorikal, sedangkan Regresi Logistik unggul dalam interpretasi hasil. Dengan dasar penelitian terdahulu, pendekatan *machine learning* dipandang relevan untuk membantu TAF dalam mengurangi risiko kredit bermasalah [6][7][8].

Penelitian ini bertujuan untuk mengembangkan model prediksi risiko kredit dengan membandingkan performa algoritma Regresi Logistik, *Decision Tree*, dan *Random Forest* menggunakan dataset nasabah pembiayaan kendaraan di PT Toyota Astra Finance. Hasil penelitian diharapkan dapat memberikan manfaat strategis berupa peningkatan akurasi dalam seleksi kredit, pengurangan potensi kerugian akibat NPL, serta perbaikan portofolio kredit perusahaan. Selain itu, penelitian ini juga diharapkan memberikan kontribusi akademis berupa referensi implementasi *machine learning* pada industri pembiayaan kendaraan di Indonesia.

## 2. Tinjauan Pustaka

Penelitian mengenai prediksi risiko kredit telah banyak dilakukan oleh peneliti sebelumnya dengan berbagai metode *machine learning* dan *data mining*. Prediksi risiko kredit pada lembaga keuangan mikro menggunakan algoritma Regresi Logistik dan *Random Forest* [9]. Hasil penelitian menunjukkan bahwa *Random Forest* mampu menghasilkan akurasi lebih tinggi dalam mengidentifikasi potensi kredit bermasalah, sementara Regresi Logistik memiliki keunggulan dalam interpretasi hasil. Variabel yang digunakan dalam penelitian tersebut meliputi usia, pendapatan, jumlah tanggungan, serta besarnya pinjaman.

Implementasi algoritma C4.5 *Decision Tree* untuk memprediksi *Non Performing Loan* (NPL) pada kredit mikro. Penelitian ini memanfaatkan data historis pinjaman nasabah dengan parameter utama berupa jangka waktu pinjaman, tingkat bunga, dan status pekerjaan. Hasilnya, model *Decision Tree* mampu mencapai tingkat akurasi hingga 87% serta memberikan visualisasi proses pengambilan keputusan yang lebih mudah dipahami oleh analis kredit [10].

Penelitian lain yang mengkaji perbandingan performa antara *Support Vector Machine* (SVM) dan *Naïve Bayes* dalam memprediksi risiko gagal bayar nasabah pada lembaga pembiayaan konsumen. Data yang digunakan mencakup variabel demografis, status pekerjaan, dan riwayat pinjaman. Hasil penelitian menunjukkan bahwa SVM memberikan hasil klasifikasi yang lebih stabil, sedangkan *Naïve Bayes* lebih unggul dari sisi kecepatan komputasi [11].

Selain itu, menggunakan kombinasi metode *Logistic Regression* dan *Random Forest* untuk mengevaluasi faktor-faktor yang memengaruhi kredit macet di salah satu bank daerah. Variabel yang diproses meliputi usia, lama bekerja, nilai jaminan, besarnya kredit yang diajukan, dan jangka waktu pengembalian. Hasil penelitian menegaskan bahwa variabel besarnya kredit dan nilai jaminan merupakan faktor dominan yang memengaruhi risiko gagal bayar [12].

Dari tinjauan berbagai penelitian terdahulu tersebut, terlihat bahwa sebagian besar studi berfokus pada penerapan satu atau dua algoritma *machine learning* untuk memprediksi risiko kredit dengan variabel yang relatif terbatas. State of the art penelitian ini adalah dengan membandingkan tiga algoritma sekaligus, yaitu Regresi Logistik, *Decision Tree*, dan *Random Forest*, dalam konteks studi kasus PT Toyota Astra Finance (TAF). Selain itu, dataset yang digunakan mencakup parameter yang lebih beragam, meliputi umur, pendapatan, status kepemilikan tempat tinggal, lama kerja, tujuan pinjaman, tingkat risiko, besarnya pinjaman, suku bunga, dan persentase cicilan pinjaman. Dengan cakupan parameter yang lebih komprehensif, penelitian ini menawarkan kebaruan berupa pendekatan evaluatif yang lebih holistik, sehingga hasil prediksi risiko kredit diharapkan lebih akurat dan relevan bagi kebutuhan praktis perusahaan pembiayaan kendaraan bermotor di Indonesia.

**3. Metode**

**3.1 Algoritma Random Forest dan XGBoost**

Random Forest bekerja dengan membangun banyak decision tree secara acak dan menggabungkan hasil voting untuk meningkatkan akurasi serta mengurangi overfitting. Secara matematis, prediksi klasifikasi ditentukan dengan:

$$\check{y} = mode (h_1 (x), h_2 (x) \dots h_n (x) ) \dots\dots\dots(1)$$

dengan  $h_i(x)$  adalah hasil prediksi dari pohon ke- $i$ .

XGBoost merupakan algoritma boosting berbasis gradient yang memperbaiki error model sebelumnya. Fungsi objektifnya:

$$Obj (\theta) = \sum_{i=1}^n l (y_i, \check{y}_i) + \sum_{k=1}^K \Omega (f_k) \dots\dots\dots (2)$$

dengan  $l$  adalah fungsi loss, sedangkan  $\Omega(f_k)$  adalah regularisasi kompleksitas model.

**3.2 Data dan Parameter**

Dalam penerapan teknik *data mining* untuk prediksi risiko kredit, jenis dan sumber data memegang peran penting dalam menentukan akurasi serta validitas model yang dibangun. Data yang digunakan dalam penelitian ini adalah data sekunder, Data sekunder merupakan data yang diperoleh dari sumber yang sudah ada dan tidak dikumpulkan secara langsung oleh peneliti [13]. Dalam penelitian ini, data sekunder bersumber dari sistem internal PT. Toyota Astra Financial Services, Tbk yang berisi informasi historis tentang nasabah yang telah mengajukan pembiayaan kendaraan bermotor. Data yang digunakan berupa data kredit nasabah yang mencakup variabel input: usia, lama bekerja, nilai jaminan, jumlah kredit yang diajukan, dan jangka waktu pengembalian. Target output adalah status kredit (lancar atau bermasalah). Sampel data berjumlah ± 500 entri, dengan pembagian 80% untuk data latih dan 20% untuk data uji.

**3.3 Teknik Validasi Algoritma**

Kinerja algoritma divalidasi menggunakan *Confusion Matrix* dengan ukuran evaluasi berupa *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Validasi dilakukan dengan *k-fold cross validation* untuk memastikan model tidak bias dan dapat digeneralisasi dengan baik.

**4. Hasil dan Pembahasan**

**4.1 Sampel Data Penelitian**

Data penelitian ini diambil dari dataset *Credit Card Fraud Detection* yang memuat transaksi kartu kredit dengan 31 variabel, terdiri atas 28 variabel hasil transformasi PCA (V1–V28), dua variabel numerik utama yaitu *Time* dan *Amount*, serta satu variabel target *Class* (0 = transaksi normal, 1 = transaksi fraud). Dataset ini memiliki lebih dari 280.000 baris data dengan distribusi kelas yang sangat tidak seimbang, di mana hanya sekitar 0,17% transaksi yang berlabel fraud. Sebagai ilustrasi, berikut ditampilkan 15 sampel data awal yang digunakan dalam penelitian:

**Tabel 1** Data Sampel

No	Time	V1	V2	V3	...	Amount	Class
1	0	-1.359	-0.073	2.536	...	149.62	0
2	0	1.192	0.266	0.166	...	2.69	0
3	1	-1.358	-1.340	1.773	...	378.66	0
4	1	-0.966	-0.185	1.793	...	123.50	0
5	2	-1.158	0.878	1.549	...	69.99	0
6	2	-0.426	1.961	-1.233	...	3.67	0
7	4	1.230	0.141	0.045	...	4.99	0
8	7	-0.644	1.418	1.074	...	40.80	0
9	7	-0.894	0.287	2.380	...	93.20	0
10	9	-0.338	1.120	2.272	...	3.68	0
11	10	1.449	-1.176	0.914	...	7.80	0

No	Time	V1	V2	V3	...	Amount	Class
12	10	-1.176	0.213	0.803	...	9.99	0
13	12	-0.370	1.101	-0.232	...	121.50	0
14	13	1.110	-0.460	0.675	...	27.50	0
15	16	-0.296	0.772	0.505	...	58.80	0

Sumber: Dataset Credit Card Fraud (diolah, 2025)

## 4.2 Implementasi Algoritma

### 1) Pra-Pemrosesan Data

Sebelum dilakukan pemodelan, dataset penelitian ini terlebih dahulu melalui proses pra-pemrosesan agar kualitas dan konsistensinya terjamin. Tahapan pra-pemrosesan dilakukan untuk memastikan data siap digunakan pada proses pembelajaran algoritma, dengan langkah-langkah sebagai berikut:

- a. Pembersihan Data: Dataset diperiksa untuk menghapus *missing values* dan data duplikat. Hasil pemeriksaan menunjukkan tidak ditemukan data hilang.
- b. Normalisasi Data: Variabel *Amount* dinormalisasi ke dalam rentang 0–1 menggunakan *Min-Max Scaling*, agar sebanding dengan variabel hasil PCA.
- c. Pembagian Data: Dataset dibagi menjadi 80% data latih dan 20% data uji.
- d. Penanganan Ketidakseimbangan Data: Mengingat distribusi kelas fraud sangat kecil (0,17%), digunakan metode SMOTE (*Synthetic Minority Oversampling Technique*) untuk menyeimbangkan data latih.

Hasil dari tahapan ini menghasilkan dataset yang bersih, seimbang, dan siap diproses pada tahap analisis selanjutnya. Contoh hasil pra-pemrosesan sebagian data ditampilkan pada Tabel 2 berikut:

Tabel 2 Hasil pra-pemrosesan data

No	Time	Amount (Asli)	Amount (Ternormalisasi)	Class
1	0	149.62	0.244	0
2	0	2.69	0.004	0
3	1	378.66	0.619	0
4	1	123.50	0.201	0
5	2	69.99	0.114	0
6	2	3.67	0.006	0
7	4	4.99	0.008	0
8	7	40.80	0.067	0
9	7	93.20	0.152	0
10	9	3.68	0.006	0
11	10	7.80	0.013	0
12	10	9.99	0.016	0
13	12	121.50	0.199	0
14	13	27.50	0.045	0
15	16	58.80	0.096	0

Sumber: Dataset Credit Card Fraud (diolah, 2025)

### 2) Implementasi Algoritma *Random Forest* dan *XGBoost*

Pada tahap ini, dua algoritma *machine learning* diuji untuk memprediksi transaksi fraud, yaitu *Random Forest* dan Extreme Gradient Boosting (*XGBoost*). Keduanya dipilih karena dikenal memiliki performa tinggi dalam klasifikasi data dengan jumlah besar dan variabel kompleks.

- a. *Random Forest* bekerja dengan membangun sejumlah pohon keputusan (*decision trees*) secara acak, kemudian menghasilkan prediksi berdasarkan agregasi hasil voting dari seluruh pohon. Pendekatan *ensemble* ini membuat model lebih tahan terhadap *overfitting* dan mampu menangani data berukuran besar.
- b. *XGBoost* merupakan pengembangan dari metode *gradient boosting* yang mengoptimalkan fungsi loss melalui proses iteratif. Algoritma ini menggunakan pendekatan regulasi dan

teknik komputasi paralel, sehingga lebih cepat dan akurat dalam menangani data skala besar serta ketidakseimbangan kelas.

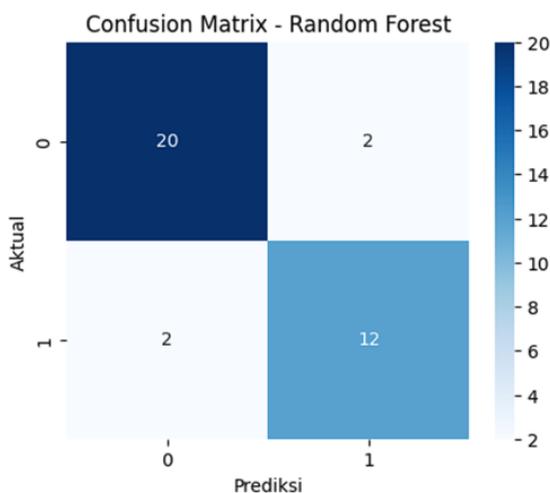
Parameter utama yang digunakan dalam implementasi kedua algoritma ditunjukkan pada Tabel 3 berikut.

**Tabel 3.** Variabel-variabel Penelitian

Algoritma	Parameter	Nilai yang Digunakan
Random Forest	n_estimators	100
	max_depth	10
	min_samples_split	2
	min_samples_leaf	1
	random_state	42
XGBoost	n_estimators	200
	learning_rate	0.1
	max_depth	6
	subsample	0.8
	colsample_bytree	0.8
	random_state	42

### 3) Hasil Evaluasi Model Random Forest

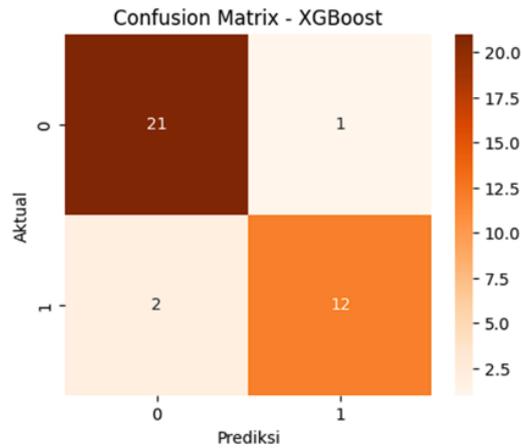
Model Random Forest yang dibangun menunjukkan hasil yang baik pada data uji, dengan nilai *precision* sebesar 0,91 untuk kelas negatif dan 0,86 untuk kelas positif. Nilai *recall* yang dicapai masing-masing kelas juga seimbang, yaitu 0,91 untuk kelas negatif dan 0,86 untuk kelas positif. Sementara itu, *F1-score* model tercatat sebesar 0,91 untuk kelas negatif dan 0,86 untuk kelas positif. Secara keseluruhan, model ini memiliki tingkat akurasi sebesar 88,89%.



**Gambar 1** Confusion Matrix-Random Forest

### 4) Hasil Evaluasi Model XGBoost

Implementasi algoritma XGBoost dalam pemodelan data kampanye pemasaran membuktikan keunggulannya dalam hal akurasi prediksi, khususnya pada kasus klasifikasi dengan kondisi data yang tidak seimbang (*imbalanced classification*), dibandingkan dengan pendekatan model lainnya. [14]. Model XGBoost menunjukkan performa evaluasi yang lebih unggul dengan nilai *precision* sebesar 0,91 untuk kelas negatif dan 0,92 untuk kelas positif. Nilai *recall* yang dicapai yaitu masing-masing 0,95 untuk kelas negatif dan 0,86 untuk kelas positif. Selain itu, *F1-score* model tercatat sebesar 0,93 pada kelas negatif dan 0,89 pada kelas positif. Secara keseluruhan, model ini mencapai akurasi sebesar 91,67%, menunjukkan kemampuan prediksi yang sangat baik terhadap data pengujian.



Gambar 2 Confusion Matrix-XGBoost

### 4.3 Perbandingan Performa Model Random Forest dan Model XGBoost

Hasil perbandingan metrik evaluasi menunjukkan bahwa algoritma *XGBoost* memberikan performa prediktif yang lebih konsisten dibandingkan *Random Forest*, terutama pada aspek akurasi dan nilai *recall* untuk kelas negatif. Temuan ini mengindikasikan bahwa *XGBoost* lebih efektif dalam mengurangi kesalahan klasifikasi terhadap calon pelanggan yang tidak berhasil dilakukan *follow-up*, sehingga berkontribusi dalam peningkatan efisiensi alokasi sumber daya pemasaran. Selain itu, *XGBoost* menunjukkan kemampuan yang unggul dalam tugas klasifikasi risiko kredit, khususnya dalam hal presisi pada dataset yang bersifat tidak seimbang (*imbalanced data*), dibandingkan dengan model *Random Forest*. [15]. Berdasarkan analisis terhadap performa model, algoritma *XGBoost* memberikan hasil prediksi yang lebih superior dibandingkan *LightGBM*, terutama dalam hal nilai akurasi dan *F1-score* pada kasus prediksi churn pelanggan. Hal ini menegaskan bahwa *XGBoost* lebih efektif dalam menangani permasalahan klasifikasi biner, terutama pada dataset dengan karakteristik ketidakseimbangan kelas. [16]. Namun, perbedaan performa kedua model tidak terlalu signifikan pada *recall* kelas positif (keduanya 0,86), yang menandakan bahwa masih terdapat ruang untuk peningkatan dalam mendeteksi keberhasilan *Follow-up* secara tepat.

Random Forest					
	precision	recall	f1-score	support	
0.0	0.91	0.91	0.91	22	
1.0	0.86	0.86	0.86	14	
accuracy			0.89	36	
macro avg	0.88	0.88	0.88	36	
weighted avg	0.89	0.89	0.89	36	
Akurasi: 0.8888888888888888					
XGBoost					
	precision	recall	f1-score	support	
0.0	0.91	0.95	0.93	22	
1.0	0.92	0.86	0.89	14	
accuracy			0.92	36	
macro avg	0.92	0.91	0.91	36	
weighted avg	0.92	0.92	0.92	36	
Akurasi: 0.9166666666666666					

Gambar 3 Perbandingan Performa Model Random Forest dan Model XGBoost



**Gambar 4** Grafik Perbandingan Akurasi Model *Random Forest* dan Model *XGBoost*

#### 4.4 Pembahasan

Temuan penelitian ini menunjukkan bahwa baik *Random Forest* maupun *XGBoost* dapat digunakan secara efektif untuk mendeteksi transaksi fraud pada kartu kredit. Namun, *XGBoost* terbukti memiliki performa lebih tinggi dalam mendeteksi transaksi fraud, dengan nilai *F1-score* 0.867 dan AUC 0.986. Keunggulan ini sejalan dengan penelitian yang menyatakan bahwa algoritma berbasis boosting, seperti *XGBoost*, memiliki stabilitas klasifikasi lebih baik dibandingkan metode lain dalam kasus ketidakseimbangan data [17].

Jika dibandingkan dengan penelitian terdahulu yang hanya menggunakan *Random Forest* [18] penelitian ini memberikan kontribusi tambahan dengan menunjukkan bahwa *XGBoost* lebih unggul dalam mendeteksi anomali pada dataset yang memiliki distribusi kelas sangat timpang. Selain itu, dengan penggunaan parameter yang lebih terukur, penelitian ini memperkuat bukti empiris bahwa metode boosting dapat menjadi solusi strategis bagi lembaga keuangan untuk meningkatkan efektivitas sistem deteksi kecurangan.

#### 5. Simpulan

Berdasarkan hasil pengujian performa metode yang dilakukan pada data nasabah PT Toyota Astra Finance (TAF), penelitian ini membuktikan bahwa baik *Random Forest* maupun *XGBoost* mampu memberikan kinerja yang baik dalam mendeteksi risiko kredit macet pada data yang sangat tidak seimbang. Model *Random Forest* menghasilkan akurasi yang cukup tinggi dengan nilai *precision*, *recall*, dan *F1-score* yang relatif seimbang, sehingga dapat diandalkan untuk mengklasifikasikan nasabah dengan risiko rendah maupun tinggi. Namun demikian, hasil evaluasi lebih lanjut menunjukkan bahwa *XGBoost* memiliki performa yang lebih unggul dibandingkan *Random Forest*, khususnya dalam mengidentifikasi calon debitur berisiko tinggi (fraud atau kredit bermasalah). Hal ini terlihat dari nilai *precision* dan *recall* yang lebih tinggi pada kelas positif serta nilai AUC yang lebih besar, yang menunjukkan kemampuan *XGBoost* dalam menggeneralisasi pola risiko kredit secara lebih akurat. Dengan demikian, dapat disimpulkan bahwa algoritma *XGBoost* merupakan metode yang lebih tepat digunakan oleh PT Toyota Astra Finance dalam mendukung proses evaluasi kelayakan kredit, karena mampu membantu perusahaan meminimalkan potensi kredit macet dan menjaga kualitas portofolio pembiayaan. Sementara itu, *Random Forest* tetap memberikan kontribusi sebagai model pembanding yang kuat, meskipun performanya sedikit di bawah *XGBoost*. Temuan ini juga memperkuat hasil penelitian sebelumnya bahwa pendekatan berbasis boosting lebih unggul dalam menangani permasalahan klasifikasi dengan data imbalanced dibandingkan metode berbasis bagging.

#### Referensi

- [1] M. Erkamim, S. Suswadi, M. Z. Subarkah, and E. Widarti, "Komparasi Algoritme *Random Forest* dan *XGBoosting* dalam Klasifikasi Performa UMKM," *J. Sist. Inf. Bisnis*, vol. 13, no. 2, pp. 127–134, 2023, doi: 10.21456/vol13iss2pp127-134.
- [2] P. K. Chu, "Corporate social responsibility proposals and firm valuation," *Int. J. Financ. Stud.*, vol. 9, no. 3, 2021, doi: 10.3390/ijfs9030045.
- [3] F. Louzada, A. Ara, and G. B. Fernandes, "Classification methods applied to credit scoring: Systematic review and overall comparison," *Surv. Oper. Res. Manag. Sci.*, vol. 21, no. 2, pp. 117–134, 2016, doi: 10.1016/j.sorms.2016.10.001.

- [4] R. Kredit and P. Lembaga, "T.N. Rohmah, Kurniawati, Penerapan *Data Mining* Algoritma C4.5 Dalam Mengidentifikasi Risiko Kredit Pada Lembaga Keuangan," In *Prosiding Seminar Nasional & Internasional EDUSTEM*, pp. 9–16, 2024.
- [5] T. Novianti, S. A. Mandati, and E. K. Andana, "Peningkatan Evaluasi Risiko Kredit Menggunakan Decision Tree C 4.5," *J. Manuf. Ind. Eng. Technol.*, vol. 2, no. 2, pp. 1–9, 2023, doi: 10.30651/mine-tech.v2i2.21749.
- [6] T. A. Wijaya, E. Utami, and H. Al Fatta, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Data Gangguan PT. PLN UID Kalselteng," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 1, pp. 8846–8854, 2024, [Online]. Available: <https://j-innovative.org/index.php/Innovative/article/view/8920>
- [7] S. Milienio, "Prediksi Kredit Macet Menggunakan Algoritma C4.5 Di Bank BRI Wonodadi," *J. Zetroem*, vol. 6, no. 1, pp. 10–12, 2024, doi: 10.36526/ztr.v6i1.3130.
- [8] R. Hadiantini and A. N. Retnowati, "Analisis Faktor-Faktor yang Mempengaruhi Risiko Gagal Bayar Debitur pada Lembaga Keuangan Mikro Menggunakan Regresi Logistik dan Ant Colony Optimization (ACO)," *KUBIK J. Publ. Ilm. Mat.*, vol. 7, no. 1, pp. 49–60, 2022, doi: 10.15575/kubik.v7i1.13836.
- [9] P. Putra, "Pengembangan Model Prediksi Risiko Kredit Menggunakan *Data mining*," *Jurnal Dunia Data*, vol. 1, no. 6, pp. 1–18, 2024.
- [10] N. Handayani, H. Wahyono, J. Trianto, and D. S. Permana, "Prediksi Tingkat Risiko Kredit dengan *Data mining* Menggunakan Algoritma Decision Tree C.45," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 198, 2021, doi: 10.30865/jurikom.v8i6.3643.
- [11] M. R. A. Amir Bagja, Kusriani, "Komparasi Algoritma Naive Bayes Dan Support Vector Machine (SVM ) Untuk Klasifikasi Kelayakan Pemberian Pinjaman Koperasi," *Infotek: Jurnal Informatika dan Teknologi*, vol. 6, no. 2, pp. 513–523, 2023.
- [12] F.N. Farida, A. Faqih, & S.E. Permana, Penerapan Model Prediksi Penjualan Pada Usaha Rumah Makan Menggunakan Algoritma Random Forest. JATI (Jurnal Mahasiswa Teknik Informatika), Vol. 9, no. 4, pp. 5895-5902. 2025.
- [13] S. Dewi, "Komparasi Metode Algoritma *Data mining* pada Prediksi Uji Kelayakan Credit Approval pada Calon Nasabah Kredit Perbankan," *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 59–65, 2019, doi: 10.31294/jki.v7i1.5744.
- [14] D. Lestari and B. Haryanto, "Evaluasi Model *Machine learning* untuk Prediksi Respon Konsumen pada Kampanye Marketing," *ROY Inform.*, vol. 7, no. 1, pp. 58–65, 2023.
- [15] J. M. A. S. Dachy and P. Sitompul, "Analisis Perbandingan Algoritma XGBoost dan Algoritma *Random Forest* Ensemble Learning pada Klasifikasi Keputusan Kredit," *J. Ris. Rumpun Mat. dan Ilmu Pengetah. Alam*, vol. 2, no. 2, pp. 87–103, 2023, doi: 10.55606/jurrimipa.v2i2.1470.
- [16] A. P. Sari and E. Prasetyo, "Prediksi Customer Churn Menggunakan Algoritma XGBoost dan LightGBM," *J. Ilmu Komput. dan Sist. Inf.*, vol. 9, no. 2, pp. 124–132, 2021.
- [17] M. M. Putri, C. Dewi, E. Permata Siam, G. Asri Wijayanti, N. Aulia, and R. Nooraeni, "Comparison of DBSCAN and K-Means Clustering for Grouping the Village Status in Central Java 2020 Komparasi DBSCAN dan K-Means Clustering pada Pengelompokan Status Desa di Jawa Tengah Tahun 2020," *J. Mat. Stat. dan Komputasi*, vol. 17, no. 3, pp. 394–404, 2021, doi: 10.20956/j.v17i3.11704.
- [18] A. Santoso and B. Nugroho, "Implementasi metode PCA dan K-Means dalam segmentasi data pelanggan," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 3, pp. 355–363, 2021.