

Analisis Komparatif Model *Logistic Regression* dan *Decision Tree* dalam Memprediksi Risiko Kredit

Alfi Khairiati^{*1}, Linda Nurhaliza², Annisa Widyadhana³

Program Studi S1 Matematika, Fakultas Sains Terapan dan Teknologi, Institut Sains Dan Teknologi Nasional, Jakarta, Jl. Moh. Kahfi II, Jagakarsa, Jakarta Selatan, Indonesia 12640

e-mail: ^{*1}alfi@istn.ac.id, ²lindanurhaliza90@gmail.com, ³annisawdbogor@gmail.com

Abstrak

Risiko kredit merupakan salah satu permasalahan utama dalam pemberian pinjaman karena berpotensi menimbulkan kredit macet yang dapat merugikan lembaga keuangan. Oleh karena itu, diperlukan model prediksi risiko kredit yang akurat dan mudah diinterpretasikan. Penelitian ini bertujuan membandingkan kinerja algoritma *Logistic Regression* dan *Decision Tree* dalam memprediksi risiko kredit debitur. Penelitian menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis machine learning menggunakan data sekunder German Credit Dataset. Sebanyak 50 observasi digunakan sebagai sampel penelitian dan dibagi menjadi data latih sebesar 80% serta data uji sebesar 20%. Tahapan penelitian meliputi prapemrosesan data, pembangunan model klasifikasi, serta evaluasi menggunakan confusion matrix, accuracy, precision, recall, F1-score, dan AUC-ROC. Hasil penelitian menunjukkan bahwa kedua model menghasilkan nilai accuracy, precision, recall, dan F1-score yang sama, namun *Logistic Regression* memperoleh nilai AUC-ROC lebih tinggi sebesar 1,000 dibandingkan *Decision Tree* sebesar 0,833. Hasil tersebut menunjukkan bahwa *Logistic Regression* memiliki kemampuan diskriminasi yang lebih baik dan lebih sesuai digunakan untuk prediksi risiko kredit pada dataset penelitian.

Kata kunci: Risiko Kredit, *Logistic Regression*, *Decision Tree*, Credit Scoring, Klasifikasi.

Abstract

Credit risk is one of the major challenges in lending activities due to its potential to generate non-performing loans and financial losses for financial institutions. Therefore, an accurate and interpretable credit risk prediction model is required. This study aims to compare the performance of *Logistic Regression* and *Decision Tree* algorithms in predicting borrowers' credit risk. The research employed a quantitative approach with an experimental method based on machine learning using secondary data from the German Credit Dataset. A total of 50 observations were used as the research sample and divided into 80% training data and 20% testing data. The research stages included data preprocessing, classification model development, and evaluation using confusion matrix, accuracy, precision, recall, F1-score, and AUC-ROC. The results showed that both models achieved identical values for accuracy, precision, recall, and F1-score, while *Logistic Regression* obtained a higher AUC-ROC value of 1.000 compared to *Decision Tree* with 0.833. These findings indicate that *Logistic Regression* has better discrimination ability and is more suitable for credit risk prediction in the dataset used.

Keywords: Credit Risk, *Logistic Regression*, *Decision Tree*, Credit Scoring, Classification.

1. Pendahuluan

Risiko kredit merupakan kemungkinan terjadinya kegagalan debitur dalam memenuhi kewajiban pembayaran pinjaman sesuai dengan perjanjian, yang berpotensi menyebabkan terjadinya kredit macet dan menimbulkan kerugian bagi lembaga keuangan. Risiko ini muncul ketika debitur tidak mampu memenuhi kewajiban pembayaran sesuai dengan perjanjian yang telah disepakati, sehingga berpotensi menimbulkan kerugian finansial dan menurunkan stabilitas lembaga keuangan (Hozeng & Aisa, 2016). Seiring dengan meningkatnya kebutuhan pembiayaan di sektor perbankan, koperasi, dan *financial technology*, kemampuan untuk menilai risiko kredit secara akurat menjadi semakin penting dalam mendukung keberlanjutan operasional lembaga keuangan.

Dalam praktiknya, penilaian risiko kredit masih banyak dilakukan menggunakan metode tradisional, seperti evaluasi manual oleh analis kredit atau sistem penilaian berbasis aturan tertentu. Pendekatan tersebut memiliki keterbatasan dalam mengolah data berjumlah besar dan berkarakteristik kompleks, serta cenderung bersifat statis dan kurang adaptif terhadap perubahan pola perilaku debitur. Kondisi ini dapat menyebabkan ketidaktepatan dalam mengidentifikasi calon debitur berisiko tinggi dan meningkatkan potensi terjadinya kredit bermasalah.

Perkembangan teknologi informasi dan ketersediaan data historis yang semakin melimpah mendorong penerapan pendekatan berbasis data mining dan *machine learning* dalam analisis risiko kredit. Data mining merupakan salah satu tahapan dalam proses *Knowledge Discovery in Databases* (KDD) yang mencakup

pembersihan data, integrasi data, pemilihan data, transformasi data, penambangan data, evaluasi pola, serta penyajian pengetahuan (Lestari et al., 2021). Pendekatan ini memungkinkan pemanfaatan data secara sistematis untuk menghasilkan prediksi risiko kredit yang lebih objektif dan konsisten. Berbagai penelitian menunjukkan bahwa pendekatan machine learning mampu meningkatkan efektivitas credit scoring melalui pemanfaatan pola data yang lebih kompleks dibandingkan metode konvensional (El Ghoujdami et al., 2024; Hafid & Khikmah, 2024; Lu et al., 2020).

Logistic Regression dan *Decision Tree* merupakan dua algoritma yang paling sering digunakan dalam pengembangan model credit scoring karena masing-masing memiliki keunggulan dalam hal akurasi, interpretabilitas, serta kemampuan menangani karakteristik data kredit yang beragam (Teary, 2025). *Logistic Regression* merupakan metode analisis prediktif yang banyak digunakan dalam klasifikasi biner karena kesederhanaannya dan kemampuannya memodelkan hubungan linier antara variabel independen dan probabilitas kejadian suatu kelas (Adrian & Verawati, 2025). Penelitian sebelumnya menunjukkan bahwa performa *Logistic Regression* dapat ditingkatkan dengan memanfaatkan pola nonlinier yang diekstraksi dari struktur *Decision Tree* tanpa mengurangi interpretabilitas model (Bensic et al., 2018). Selain itu, penerapan *Logistic Regression* pada kasus pemberian Kredit Usaha Rakyat menunjukkan kemampuan klasifikasi yang tinggi dalam mendukung pengambilan keputusan kredit (Safitri et al., 2024). Sementara itu, *Decision Tree* merupakan metode klasifikasi dalam data mining yang memanfaatkan data untuk menemukan hubungan antara variabel

target dan variabel prediktor melalui pembentukan struktur pohon keputusan (Syafii et al., 2024).

Beberapa penelitian terdahulu menunjukkan hasil yang berbeda terkait performa metode klasifikasi pada prediksi risiko kredit sehingga evaluasi model masih perlu dilakukan pada dataset dan konfigurasi yang berbeda (El Ghoujdam et al., 2024; Hafid & Khikmah, 2024; Lu et al., 2020). Berdasarkan latar belakang tersebut, penelitian ini bertujuan membandingkan kinerja *Logistic Regression* dan *Decision Tree* dalam memprediksi risiko kredit menggunakan *German Credit Dataset*. Penelitian ini tidak hanya mengevaluasi kinerja klasifikasi kedua metode, tetapi juga mempertimbangkan aspek interpretabilitas model melalui analisis koefisien regresi dan struktur pohon keputusan. Hasil penelitian diharapkan memberikan kontribusi empiris sebagai referensi dalam pemilihan model *credit scoring* yang akurat dan mudah diinterpretasikan pada lembaga keuangan.

2. Metodologi

Metodologi penelitian ini dirancang untuk membandingkan kinerja dua algoritma klasifikasi, yaitu *Logistic Regression* dan *Decision Tree*, dalam memprediksi risiko kredit debitur. Pendekatan yang digunakan bersifat kuantitatif dengan eksperimen komputasional berbasis data mining dan machine learning. Seluruh tahapan penelitian disusun secara sistematis agar hasil yang diperoleh dapat dianalisis secara objektif dan direplikasi.

2.1 Jenis Dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen. Pendekatan ini digunakan untuk menguji kinerja model klasifikasi berdasarkan data

historis risiko kredit. Objek penelitian difokuskan pada proses prediksi status kredit debitur ke dalam dua kategori, yaitu debitur lancar dan debitur gagal bayar. Eksperimen dilakukan dengan membangun model *Logistic Regression* dan *Decision Tree* menggunakan data latih, kemudian menguji performa model menggunakan data uji. Hasil prediksi dari masing-masing model selanjutnya dievaluasi dan dibandingkan menggunakan metrik evaluasi yang relevan dalam konteks risiko kredit.

2.2 Sumber dan Karakteristik Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari dataset publik risiko kredit, yaitu *German Credit Data*. Dataset tersebut bersifat *cross-sectional* dan merepresentasikan karakteristik debitur pada satu periode pengamatan tertentu. Variabel target dalam penelitian ini adalah status risiko kredit yang diklasifikasikan ke dalam dua kelas, yaitu 0 untuk debitur dengan status kredit lancar dan 1 untuk debitur dengan status gagal bayar. Variabel prediktor yang digunakan mencakup karakteristik demografis dan finansial debitur, yang terdiri atas usia dan jenis kelamin, jenis pekerjaan, status tempat tinggal, tingkat kepemilikan tabungan, saldo rekening giro, jumlah kredit yang diajukan, lama pinjaman, serta tujuan pengajuan kredit. Kombinasi variabel numerik dan kategorikal tersebut memungkinkan penerapan metode *Logistic Regression* dan *Decision Tree* untuk menganalisis pola dan faktor yang mempengaruhi risiko kredit.

Tabel 1. Variabel Penelitian

No.	Nama Variabel	Jenis Data
1.	Usia	Numerik
2.	Jenis Kelamin	Kategorikal

3.	Jenis Pekerjaan	Ordinal
4.	Status Tempat Tinggal	Kategorikal
5.	Rekening Tabungan	Ordinal
6.	Rekening Giro	Numerik
7.	Jumlah Kredit	Numerik
8.	Lama Pinjaman	Numerik
9.	Tujuan Kredit	Kategorikal

2.3 Tahapan Penelitian

Tahapan penelitian dilakukan secara berurutan untuk memastikan kualitas data dan keandalan model yang dihasilkan. Alur penelitian dimulai dari pengumpulan data hingga analisis hasil pemodelan. Tahap pertama adalah pengumpulan data dari sumber terbuka yang menyediakan data risiko kredit. Data yang diperoleh kemudian dianalisis secara awal untuk memahami struktur, jenis variabel, dan distribusi kelas.

Tahap kedua adalah preprocessing data yang mencakup pemeriksaan nilai hilang, pengkodean variabel kategorikal, serta normalisasi variabel numerik. Normalisasi diterapkan khusus untuk model *Logistic Regression* agar setiap fitur berada pada skala yang sebanding. Selain itu, dilakukan penanganan ketidakseimbangan kelas untuk meningkatkan kemampuan model dalam mendeteksi debitur gagal bayar. Tahap ketiga adalah pembagian data menjadi data latih dan data uji dengan proporsi 70–80% untuk data latih dan 20–30% untuk data uji. Pembagian ini bertujuan untuk menghindari *overfitting* dan memperoleh estimasi kinerja model yang lebih objektif.

Tahap keempat adalah pembangunan model klasifikasi. Model *Logistic Regression* dibangun untuk menganalisis pengaruh variabel prediktor terhadap probabilitas gagal bayar, sedangkan model *Decision Tree* dibangun untuk menangkap hubungan

nonlinier antarvariabel melalui struktur pohon keputusan. Tahap terakhir adalah evaluasi dan perbandingan kinerja model. Hasil prediksi dari kedua model dianalisis menggunakan metrik evaluasi dan dibandingkan untuk menentukan model yang memiliki performa lebih baik dalam konteks risiko kredit.

2.4 Metode Pemodelan

Penelitian ini menggunakan dua algoritma klasifikasi, yaitu *Logistic Regression* dan *Decision Tree*, untuk memprediksi risiko kredit debitur. Kedua model dipilih karena memiliki karakteristik yang berbeda dalam memodelkan hubungan antara variabel prediktor dan variabel target, sehingga memungkinkan dilakukan analisis komparatif baik dari sisi kinerja prediktif maupun interpretabilitas model.

Logistic Regression

Logistic Regression merupakan metode klasifikasi biner yang digunakan untuk memodelkan probabilitas terjadinya suatu peristiwa, dalam hal ini kejadian gagal bayar kredit. Model ini menghubungkan variabel prediktor dengan probabilitas kejadian melalui fungsi logit dan fungsi sigmoid. Secara matematis, probabilitas debitur mengalami gagal bayar dinyatakan sebagai berikut:

$$P(Y = 1) = \frac{1}{1 + e^{-\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}}$$

Interpretasi pengaruh variabel prediktor terhadap risiko kredit dilakukan melalui nilai koefisien regresi dan *odds ratio*. Nilai *odds ratio* diperoleh dari eksponensial koefisien regresi, yaitu e^{β_i} . Nilai *odds ratio* lebih besar dari satu menunjukkan peningkatan peluang gagal bayar, sedangkan nilai kurang dari satu menunjukkan penurunan peluang gagal bayar. Model *Logistic Regression* dilatih menggunakan

data latih yang telah melalui proses normalisasi untuk memastikan setiap variabel numerik berada pada skala yang sebanding. Pendekatan ini bertujuan untuk meningkatkan stabilitas estimasi koefisien dan kinerja model secara keseluruhan.

Decision Tree

Decision Tree merupakan algoritma klasifikasi yang membangun model prediksi dalam bentuk struktur pohon keputusan. Proses pembentukan pohon dilakukan melalui pembagian data secara rekursif berdasarkan atribut yang paling mampu memisahkan kelas target. Kriteria pemilihan atribut pada setiap node ditentukan menggunakan ukuran impuritas. Salah satu ukuran yang umum digunakan adalah *entropy*, yang dirumuskan sebagai berikut:

$$Entropy = - \sum_{i=1}^k p_i \log_2(p_i)$$

Dengan p_i menyatakan proporsi data pada kelas ke i dalam suatu himpunan data S . Pemilihan atribut terbaik dilakukan berdasarkan nilai *information gain*, yaitu selisih antara nilai entropy sebelum dan sesudah pembelahan data. Secara matematis, *information gain* dinyatakan sebagai:

$$Gain(S, A) = Entropy(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \log_2 \left(\frac{|S_v|}{|S|} \right)$$

Selain entropy, kriteria pembelahan juga dapat menggunakan *Gini impurity* yang didefinisikan sebagai:

$$Gini = 1 - \sum_{i=1}^k p_i^2$$

Nilai impuritas lebih rendah menunjukkan bahwa node semakin homogen. Dalam penelitian ini, pemilihan kriteria pembelahan dan parameter model disesuaikan untuk mengurangi risiko *overfitting*. Beberapa parameter pembatas yang digunakan antara lain kedalaman

maksimum pohon dan jumlah minimum sampel pada setiap node. *Decision Tree* tidak memerlukan asumsi linearitas maupun normalisasi data, sehingga mampu menangkap hubungan nonlinier antarvariabel. Namun demikian, kompleksitas struktur pohon dikendalikan agar model tetap memiliki kemampuan generalisasi yang baik terhadap data uji.

2.5 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan menggunakan beberapa metrik, yaitu confusion matrix, akurasi, presisi, recall, F1-score, dan AUC-ROC. Metrik tersebut umum digunakan dalam penelitian klasifikasi risiko kredit karena mampu memberikan gambaran menyeluruh terhadap kemampuan model dalam melakukan identifikasi dan diskriminasi kelas (El Ghoujdam et al., 2024; Hafid & Khikmah, 2024; Lu et al., 2020). Dalam konteks risiko kredit, metrik recall untuk kelas gagal bayar menjadi perhatian utama karena kesalahan dalam mendeteksi debitur berisiko tinggi dapat menimbulkan kerugian finansial yang signifikan. Perbandingan hasil evaluasi dari kedua model digunakan sebagai dasar untuk menganalisis keunggulan dan keterbatasan masing-masing metode dalam memprediksi risiko kredit.

3. Hasil Dan Pembahasan

3.1 Pengambilan Data

Data yang digunakan dalam penelitian ini diperoleh dari *German Credit Dataset* yang tersedia secara terbuka melalui repositori UCI *Machine Learning Repository*. Dataset asli terdiri atas sejumlah observasi yang menggambarkan karakteristik debitur dan satu variabel target berupa risiko kredit. Dalam penelitian ini digunakan sebanyak 50 observasi yang dipilih sebagai sampel

penelitian dengan mempertimbangkan kebutuhan eksperimen komparatif serta kemudahan interpretasi hasil model. Penggunaan subset data dilakukan agar proses analisis, visualisasi model, dan interpretasi hubungan antarvariabel dapat dilakukan secara lebih terfokus tanpa mengubah tujuan utama penelitian, yaitu membandingkan kinerja *Logistic Regression* dan *Decision Tree* dalam prediksi risiko kredit. Dataset ini dipilih karena banyak digunakan sebagai acuan dalam penelitian klasifikasi risiko kredit dan memiliki karakteristik yang relevan dengan permasalahan yang dikaji.

3.2 Statistik Deskriptif

Tabel 2. Statistik Deskriptif Variabel Penelitian

Variabel	Mean	Min	Max	Std Deviasi
<i>Duration</i> (bulan)	21,46	6	60	13,97
<i>Credit Amount</i>	3.491	409	12.579	2.677
<i>Age</i> (tahun)	37,98	22	67	12,23
<i>Existing Credits</i>	1,38	1	3	0,57
<i>Num Dependents</i>	1,16	1	2	0,37

Statistik deskriptif menunjukkan bahwa sebagian besar debitur berada pada usia produktif dengan durasi kredit yang bervariasi. Perbedaan nilai rata-rata dan simpangan baku pada beberapa variabel menunjukkan adanya variasi profil debitur yang menjadi dasar dalam pemodelan klasifikasi risiko kredit.

3.3 Pemeriksaan *Missing Values*

Pemeriksaan data menunjukkan tidak terdapat nilai yang hilang pada seluruh variabel.

3.4 Prapemrosesan Data

Tahap kedua adalah preprocessing data yang mencakup pemeriksaan nilai hilang, pengkodean variabel kategorikal, serta normalisasi variabel numerik. Normalisasi diterapkan khusus untuk model *Logistic Regression* agar setiap fitur berada pada skala yang sebanding. Tahap ketiga adalah pembagian data menjadi data latih dan data uji dengan proporsi 80% dan 20% menggunakan pembagian terstratifikasi. Pembagian ini bertujuan untuk menghindari *overfitting* dan memperoleh estimasi kinerja model yang lebih objektif.

3.5 Logistic Regression

Model *Logistic Regression* digunakan untuk mengklasifikasikan risiko kredit debitur ke dalam kategori berisiko dan tidak berisiko. Evaluasi model dilakukan menggunakan *confusion matrix* untuk melihat kemampuan klasifikasi, serta analisis koefisien regresi untuk mengidentifikasi pengaruh masing-masing variabel terhadap risiko kredit.

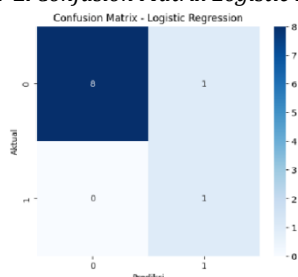
Tabel 3. Koefisien *Logistic Regression*

Index	Variabel	Koefisien
0	Status_Checking	-1.060964
1	Duration	0.102652
2	Credit_History	0.117758
3	Purpose	0.293189
4	Credit_Amount	0.016955
5	Savings	-1.049260
6	Employment	-0.326422
7	Installment_Rate	1.353462
8	Personal_Status_Sex	-0.518936
9	Guarantors	-0.303931
10	Residence_Durations	-0.656096
11	Property	0.396106
12	Age	-0.098162
13	Other_Installments	0.489963
14	Housing	0.017202
15	Existing_Credits	0.337479
16	Job	-0.085628
17	Dependents	-0.286829
18	Telephone	-0.908812

19	Foreign_Worker	-0.062355
----	----------------	-----------

Hasil estimasi koefisien menunjukkan bahwa beberapa variabel memiliki pengaruh yang berbeda terhadap risiko kredit. Variabel dengan koefisien positif cenderung meningkatkan peluang gagal bayar, sedangkan variabel dengan koefisien negatif cenderung menurunkan peluang gagal bayar. Berdasarkan hasil model, variabel *Credit Amount* memiliki koefisien positif yang menunjukkan bahwa peningkatan jumlah kredit berpotensi meningkatkan risiko gagal bayar. Sebaliknya, variabel *Age* memiliki koefisien negatif yang mengindikasikan bahwa debitur dengan usia lebih tinggi cenderung memiliki risiko kredit yang lebih rendah.

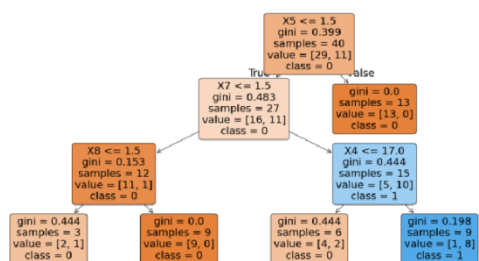
Gambar 1. *Confusion Matrix Logistic Regression*



Berdasarkan *confusion matrix*, model *Logistic Regression* berhasil mengklasifikasikan 9 dari 10 data uji dengan benar. Terdapat satu kesalahan klasifikasi pada kelas tidak berisiko yang diprediksi sebagai berisiko, sedangkan seluruh debitur berisiko pada data uji berhasil terdeteksi dengan benar. Hasil ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mengidentifikasi debitur berisiko, sehingga dapat membantu meminimalkan risiko kesalahan pemberian kredit.

3.6 Decision Tree

Gambar 2. Visualisasi *Decision Tree* pada Prediksi Risiko Kredit



Struktur pohon keputusan menunjukkan bahwa atribut pada node akar (*root node*), memiliki peran paling penting dalam proses klasifikasi risiko kredit. Selanjutnya, variabel X7, X8, dan X4 digunakan sebagai dasar pembentukan cabang-cabang keputusan berikutnya. Hal ini menunjukkan bahwa variabel-variabel tersebut memiliki kontribusi yang lebih besar dalam membedakan debitur berisiko dan tidak berisiko. Selain memberikan hasil klasifikasi,

Decision Tree juga menghasilkan aturan keputusan yang mudah dipahami dan diinterpretasikan. Namun, model ini cenderung lebih sensitif terhadap variasi data sehingga berpotensi mengalami penurunan kinerja pada data yang berbeda.

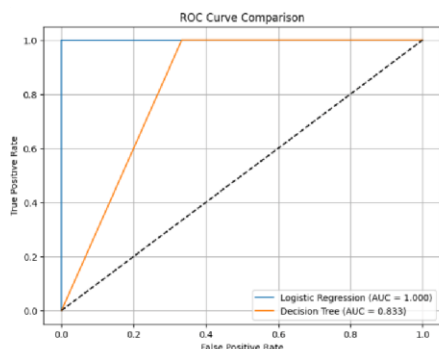
3.7 Evaluasi Kinerja Model

Tabel 4. Perbandingan Kinerja Model

Model	<i>Logistic Regression</i>	<i>Decision Tree</i>
Accuracy	0.900	0.900
Precision	0.500	0.500
Recall	1.000	1.000
F1-Score	0.677	0.677
AUC-ROC	1.000	0.833

Tabel 4 menunjukkan bahwa *Logistic Regression* dan *Decision Tree* menghasilkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang sama pada data pengujian. Kedua model mampu mengklasifikasikan 9 dari 10 data uji dengan benar dan berhasil mendeteksi seluruh debitur berisiko. Namun demikian, berdasarkan kurva ROC, *Logistic Regression* memperoleh nilai AUC sebesar 1,000, lebih tinggi dibandingkan *Decision Tree* yang memperoleh nilai AUC sebesar 0,833. Hasil ini menunjukkan bahwa *Logistic Regression* memiliki kemampuan diskriminasi yang lebih baik dalam membedakan debitur berisiko dan tidak berisiko. Oleh karena itu, *Logistic Regression* dinilai lebih unggul dan lebih sesuai digunakan untuk prediksi risiko kredit pada dataset yang digunakan dalam penelitian ini.

Gambar 3. Perbandingan Kurva ROC *Logistic Regression* dan *Decision Tree*



Berdasarkan Gambar 3, *Logistic Regression* memperoleh nilai AUC sebesar 1,000, sedangkan *Decision Tree* memperoleh nilai AUC sebesar 0,833. Nilai AUC yang lebih tinggi menunjukkan bahwa *Logistic Regression* memiliki kemampuan diskriminasi yang lebih baik dalam membedakan debitur berisiko dan tidak berisiko. Hasil ini mengindikasikan bahwa *Logistic Regression* lebih konsisten dalam menghasilkan probabilitas prediksi yang akurat dibandingkan *Decision Tree* pada dataset yang digunakan.

4. Kesimpulan

Berdasarkan hasil evaluasi, *Logistic Regression* dan *Decision Tree* menghasilkan nilai *accuracy*, *precision*, *recall*, dan F1-score yang sama pada data pengujian. Namun, *Logistic Regression* menunjukkan nilai AUC-ROC yang lebih tinggi dibandingkan *Decision Tree*, yang menunjukkan kemampuan diskriminasi yang lebih baik dalam membedakan debitur berisiko dan tidak berisiko.

Dari sisi interpretabilitas, *Logistic Regression* menunjukkan bahwa variabel *Credit Amount* dan *Installment Rate* berpengaruh positif terhadap risiko gagal bayar, sedangkan variabel *Age* berpengaruh negatif. Sementara itu, *Decision Tree* menunjukkan bahwa beberapa variabel tertentu berperan dominan dalam proses klasifikasi melalui struktur pohon

keputusan. Dengan mempertimbangkan kemampuan diskriminasi dan kemudahan interpretasi, *Logistic Regression* dinilai lebih sesuai digunakan untuk prediksi risiko kredit pada dataset yang digunakan dalam penelitian ini.

Penelitian ini memberikan kontribusi empiris dalam membandingkan kinerja dan interpretabilitas kedua metode sebagai referensi pemilihan model *credit scoring* pada lembaga keuangan. Penelitian ini masih memiliki keterbatasan pada jumlah observasi yang digunakan sehingga hasil yang diperoleh belum dapat digeneralisasikan secara luas. Penelitian selanjutnya disarankan menggunakan jumlah data yang lebih besar serta melakukan pengujian pada beberapa metode klasifikasi tambahan untuk memperoleh evaluasi yang lebih komprehensif.

Daftar Pustaka

- Adrian, A., & Vewawati, I. (2025). Analisis performa Logistic Regression dan Random Forest dalam klasifikasi kelayakan penerimaan kredit. *Jurnal Sistem Informasi dan Teknologi*, 4(2), 148–158.
- Bensic, M., Sarlija, N., & zekic-Susac, M. (2018). *Machine learning for credit scoring: Improving Logistic Regression with non-linear Decision Tree effects*. [<https://arxiv.org/abs/1802.03820>] <https://arxiv.org/abs/1802.03820>
- El Ghoujdam, M., Chaabita, R., Elkhalfi, O., & Zehraoui, K. (2024). Consumer credit risk analysis through artificial intelligence: A comparative study between the classical approach of logistic regression and advanced machine learning techniques. *Cogent Economics & Finance*, 12(1). <https://doi.org/10.1080/23322039.2>

- [024.2414926\]\(https://doi.org/10.1080/23322039.2024.2414926\)](https://doi.org/10.1080/23322039.2024.2414926)
- Hafid, N., & Khikmah, L. (2024).** Perbandingan metode algoritma C4.5, Naïve Bayes, dan Logistic Regression untuk penentuan kelayakan penerima kredit. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 14(2), 85–93.
- Hozeng, S., & Aisa, S. (2016).** Aplikasi data mining dengan menggunakan metode Decision Tree untuk prediksi penentuan risiko kredit. *Jurnal Informatika*, 5(2), 1–10.
- Lestari, W. J., Army, W. L., & Habibie, D. R. (2021).** Analisa risiko kredit macet dengan pendekatan data mining (Studi kasus: Koperasi Putra Kembar). *Jurnal Sistem Informasi*, 9(1), 58–64.
- Safitri, E., Hazirah, A., & Rahmadeni. (2024).** Estimasi pemberian pinjaman Kredit Usaha Rakyat (KUR) di Bank BRI Unit Tangkerang Pekanbaru menggunakan Logistic Regression Analysis. *Zeta-Math Journal*, 9(1), 1–9. [https://doi.org/10.31102/zeta.2024.9.1.1-9] <https://doi.org/10.31102/zeta.2024.9.1.1-9>.
- Safitri, E., Hazirah, A., & Rahmadeni. (2024).** Estimasi pemberian pinjaman Kredit Usaha Rakyat (KUR) di Bank BRI Unit Tangkerang Pekanbaru menggunakan Logistic Regression Analysis. *Zeta-Math Journal*, 9(1), 1–9. [https://doi.org/10.31102/zeta.2024.9.1.1-9] <https://doi.org/10.31102/zeta.2024.9.1.1-9>.
- Safitri, E., Ribhi, A. A., & Astutik, L. Y. (2024).** Analysis of customer repeat order predictions using the Decision Tree algorithm in transportation companies. *Journal of Information Technology and Computer Science*, 4, 1372–1378.
- Teary, M. G. (2025).** Optimalisasi model credit scoring menggunakan algoritma Logistic Regression dan Decision Tree. *Jurnal Ilmiah Informatika*, 16(3), 352–358.
- Wang, Y., Lu, Y., Zhang, Y., Yu, X., & Wang, Y. (2020).** A comparative assessment of credit risk model based on machine learning: A case study of bank loan data. *Procedia Computer Science*, 174, 141–149. [https://doi.org/10.1016/j.procs.2020.06.069] <https://doi.org/10.1016/j.procs.2020.06.069>.