

## Optimasi Kinerja Decision Tree C4.5 dengan Metode *Backward Elimination* pada Sistem Penilaian Kredit

Amrin<sup>1</sup>, Andi Diah Kuswanto<sup>2</sup>, Ahmad Fauzi<sup>3</sup>

<sup>1,2,3</sup>Fakultas Teknik dan Informatika  
<sup>1,2,3</sup>Universitas Bina Sarana Informatika

e-mail: <sup>1</sup>amrin.ain@bsi.ac.id, <sup>2</sup>andi.ahk@bsi.ac.id, <sup>3</sup>ahmad.afz@bsi.ac.id

Diterima	Direvisi	Disetujui
12-11-2025	18-12-2025	20-12-2025

**Abstrak** - Penggunaan kredit telah menjadi fenomena yang meluas di masyarakat, namun di balik popularitasnya terdapat risiko signifikan berupa peningkatan kredit macet. Masalah ini sering kali berakar pada histori pengelolaan keuangan yang buruk, di mana kegagalan nasabah untuk memenuhi kewajiban pembayaran utang secara langsung meningkatkan risiko kredit bagi lembaga pemberi pinjaman. Untuk mengatasi tantangan ini, penelitian ini mengimplementasikan dan membandingkan dua model klasifikasi untuk menilai kelayakan kredit yaitu algoritma *Decision Tree* C4.5 dan C4.5 yang dioptimasi dengan teknik seleksi fitur *Backward Elimination*. Menggunakan dataset 481 catatan kredit kendaraan yang diklasifikasikan sebagai "baik" dan "buruk", model ini dikembangkan dengan sebelas variabel independen, termasuk status tanggungan, usia, pendidikan, pekerjaan, pendapatan, dan uang muka. Hasil pengujian menunjukkan bahwa model C4.5 tanpa optimasi mencapai akurasi sebesar 91,90% dengan nilai *Area Under Curve* (AUC) 0,915. Sebaliknya, model yang mengintegrasikan *Backward Elimination* menunjukkan peningkatan kinerja yang signifikan, dengan akurasi mencapai 94,80% dan AUC sebesar 0,973. Temuan ini mengkonfirmasi bahwa penerapan optimasi *Backward Elimination* secara efektif meningkatkan kemampuan prediktif model klasifikasi kelayakan kredit.

Kata Kunci: *Decision Tree C4.5, Backward Elimination, ROC curve*

### PENDAHULUAN

Praktik pemberian kredit memiliki sejarah yang panjang dan integral dalam perkembangan peradaban ekonomi, dengan jejak awal dokumen kredit ditemukan di peradaban Mesir kuno, Babilonia, dan Asyur sekitar 3.000 tahun yang lalu (Himberg, 2020). Di era modern, kredit menjadi tulang punggung perekonomian global, namun diiringi oleh tantangan utama berupa risiko kredit. Risiko ini, yang berpotensi menyebabkan kerugian finansial signifikan, merupakan salah satu masalah krusial yang dihadapi oleh lembaga keuangan di seluruh dunia (Alfian & Nugroho, 2024).

Di era digital, evaluasi kredit menjadi semakin kompleks. Ledakan big data, yang dipicu oleh kemajuan teknologi informasi dan jaringan, memungkinkan pengumpulan data dalam volume dan kecepatan yang belum pernah terjadi sebelumnya dari berbagai sumber yang tidak terstruktur (Chern et al., 2021). Mengelola data masif ini, ditambah dengan persyaratan kredit yang dinamis, menuntut pendekatan pemrosesan yang lebih canggih. Proses analisis kredit tradisional, yang mengandalkan prinsip Character, Capital, Capacity, Collateral, dan Condition (5C), seringkali

memerlukan verifikasi lapangan yang profesional dan memakan waktu lama, terutama saat harus menangani sejumlah besar dokumen nasabah (Lasena & Ahmad, 2023). Selain itu, data penilaian kredit modern sering kali memiliki dimensi yang tinggi. Fitur yang berlebihan atau tidak relevan dapat menyebabkan masalah overfitting pada model prediksi, yang pada akhirnya menurunkan kinerja dan generalisasi model (Jin et al., 2021).

Untuk mengatasi tantangan ini, industri perbankan dan keuangan beralih ke data mining dan machine learning (Oktafriani et al., 2023). Teknik-teknik ini memungkinkan identifikasi dan prediksi kelayakan nasabah secara lebih efisien dengan menganalisis data historis pinjaman (Amrin & Pahlevi, 2022). Model klasifikasi dalam data mining, seperti Decision Tree, Support Vector Machine (SVM), dan k-Nearest Neighbor (k-NN), didukung oleh metode pra-pemrosesan seperti diskritisasi fitur, resampling data, dan terutama seleksi fitur (*feature selection*). Proses seleksi fitur sangat krusial untuk mengurangi beban komputasi, meningkatkan efisiensi, dan memperbaiki interpretasi model dengan memilih subset fitur yang paling relevan (Ziamba et al., 2021). Dengan demikian, peningkatan akurasi model prediksi kredit menjadi sangat penting tidak



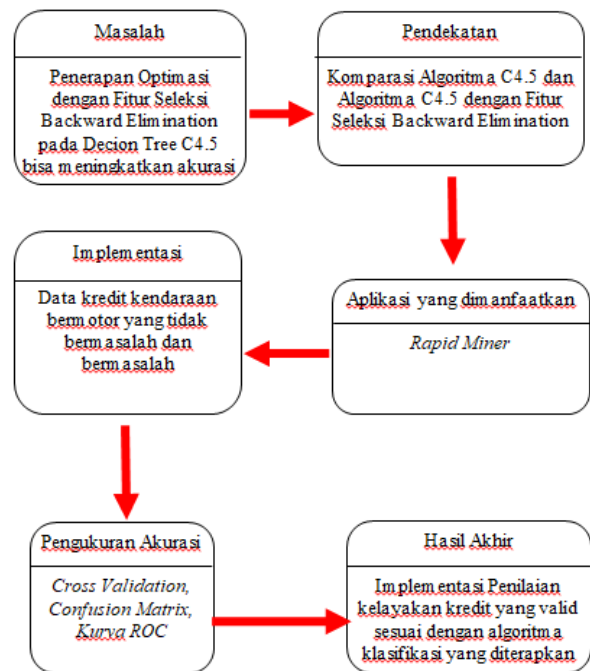
hanya untuk melindungi dana investor tetapi juga untuk mengurangi ketidakseimbangan informasi antara bisnis, lembaga keuangan, dan masyarakat luas (Song, 2023).

Sejalan dengan hal itu, berbagai penelitian telah dilakukan untuk meningkatkan akurasi model prediksi kredit. Beberapa studi berfokus pada penggunaan algoritma optimasi. Sebagai contoh, bahwa optimasi Particle Swarm Optimization (PSO) pada Decision Tree secara signifikan meningkatkan akurasi analisis risiko kredit di Bank DKI dari 97,83% menjadi 99,13% (Putry et al., 2024). Temuan serupa juga dilaporkan oleh (Novichasari, 2021) dan (Agustian & Bisri, 2019), di mana penggabungan PSO dengan teknik klasifikasi seperti SVM dan Naive Bayes berhasil meningkatkan akurasi dan presisi model. Sementara itu, penelitian lain mengeksplorasi metode seleksi fitur. (Ubaedi & Djaksana, 2022) membuktikan bahwa penggunaan Forward Selection dan Stratified Sampling mampu meningkatkan akurasi algoritma C4.5 sebesar 9,2% dalam prediksi kelayakan kredit. Di sisi lain, tidak semua metode optimasi selalu berhasil. (Religia et al., 2021) menemukan bahwa optimasi Genetic Algorithm (GA) dan Bagging belum mampu meningkatkan akurasi algoritma Random Forest untuk dataset yang tidak seimbang (*imbalance*) (Religia et al., 2021).

Berdasarkan kajian pustaka tersebut, terlihat bahwa berbagai pendekatan optimasi, baik melalui algoritma optimasi maupun teknik seleksi fitur, telah terbukti efektif dalam meningkatkan kinerja model prediksi kredit. Namun, masih ada ruang untuk eksplorasi teknik seleksi fitur lainnya. Meskipun metode seperti Forward Selection telah dieksplorasi, penerapan teknik *Backward Elimination* pada algoritma Decision Tree C4.5 secara khusus untuk konteks kelayakan kredit masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan membandingkan kinerja algoritma Decision Tree C4.5 tanpa optimasi dengan Decision Tree C4.5 yang dioptimasi menggunakan teknik seleksi fitur *Backward Elimination* dalam mengklasifikasikan status kelayakan kredit.

## METODE PENELITIAN

Penelitian ini melalui beberapa tahapan sebagaimana ditunjukkan pada kerangka pemikiran di Gambar 1. Fokus utamanya adalah menilai apakah penerapan optimasi melalui fitur seleksi *Backward Elimination* pada algoritma Decision Tree C4.5 dapat meningkatkan akurasi. Untuk menjawab pertanyaan tersebut, dibuat dua pendekatan (model), yaitu Decision Tree C4.5 tanpa seleksi fitur dan Decision Tree C4.5 yang dipadukan dengan *Backward Elimination*. Kedua model tersebut kemudian diuji kinerjanya. Proses pengujian dilakukan menggunakan metode Cross Validation, Confusion Matrix, dan kurva ROC. Dalam tahap pengembangan aplikasi berdasarkan model tersebut, digunakan perangkat lunak RapidMiner.



Sumber: Hasil Penelitian (2025)

Gambar 1. Kerangka Pemikiran Penelitian

## HASIL DAN PEMBAHASAN

### A. Analisa Data

Dataset yang digunakan dalam penelitian ini berjumlah 481 data kredit kendaraan bermotor, mencakup debitur bermasalah maupun tidak bermasalah. Penelitian ini melibatkan tiga belas variabel input, yaitu: (1) Status Perkawinan, (2) Jumlah Tanggungan, (3) Umur, (4) Status Tempat Tinggal, (5) Kepemilikan Rumah, (6) Pekerjaan, (7) Status Pekerjaan, (8) Status Perusahaan, (9) Penghasilan, (10) Uang Muka, (11) Pendidikan, (12) Lama Tinggal, dan (13) Kondisi Rumah. Gambar 1 menampilkan contoh dataset yang digunakan sebagai bahan pengujian terhadap algoritma yang diteliti, yaitu *decision tree* C4.5 dengan Metode *Backward Elimination*. *Backward Elimination* merupakan teknik seleksi fitur dalam proses klasifikasi yang

bertujuan menyederhanakan model dengan menghilangkan atribut yang kurang berpengaruh, sehingga meningkatkan efisiensi dan kinerja model menjadi lebih optimal.

Tabel 1 Sampel Dataset

status perkawinan	jumlah tanggungan	pendidikan terakhir	usia	kepemilikan rumah	lama tinggal	kondisi rumah
menikah	2-3	SLTP	21-55	ortu	>5	permanen
belum menikah	tdk ada	SLTA	21-55	ortu	1-3	permanen
menikah	2-3	SLTA	21-55	KPR	3-5	permanen
belum menikah	tdk ada	SLTA	<21-60	ortu	1-3	permanen
menikah	>3	SD	21-55	atas nama	>5	non permanen
menikah	2-3	SLTP	21-55	milik sendiri	>5	permanen
menikah	2-3	SLTA	21-55	milik sendiri	>5	permanen

## B. Pengujian Model

Penelitian ini dilakukan dengan eksperimen pengujian pada model yang diusulkan. Kemudian dilakukan evaluasi dan validasi model untuk menghasilkan nilai accuracy dan AUC. Pengujian dilakukan menggunakan RapidMiner dengan menerapkan operator 10-fold cross-validation untuk memperoleh nilai akurasi dan AUC dari masing-masing algoritma yang diuji. Evaluasi kinerja model dilakukan melalui *Confusion Matrix* serta *ROC Curve* atau *Area Under the Curve (AUC)*.

### 1. Confusion Matrix

#### a. Algoritma Decision Tree C4.5

Tabel 2 menampilkan confusion matrix untuk algoritma Decision Tree C4.5. Dari tabel tersebut terlihat bahwa 152 data dengan label “bad” berhasil diprediksi dengan benar, sedangkan 27 data yang sebenarnya “bad” justru diprediksi sebagai “good”. Sementara itu, 290 data berlabel “good” diklasifikasikan secara tepat, dan 12 data yang sebenarnya “good” diprediksi sebagai “bad”.

Tabel 2. Model *Confusion Matrix* untuk Decision Tree C4.5

accuracy: 91.80% +/- 4.50% (mikro: 91.88%)			
	true bad	true good	class precision
pred. bad	152	12	92.88%
pred. good	27	290	91.48%
class recall	84.02%	98.03%	

Sumber: Hasil Pengolahan Menggunakan RapidMiner (2025)

#### b. Algoritma Decision Tree C4.5 dengan *Backward Elimination*

Tabel 3 menyajikan confusion matrix untuk algoritma Decision Tree C4.5 yang telah diterapkan dengan *Backward Elimination*. Pada tabel tersebut terlihat bahwa 298 data berlabel “good” berhasil diprediksi dengan benar, sementara 4 data yang sebenarnya “good” justru terklasifikasi sebagai “bad”. Selain itu, 158 data berlabel “bad” diprediksi secara tepat, dan 21 data yang seharusnya “bad” diprediksi sebagai “good”.

Tabel 3. Model *Confusion Matrix* untuk Metode *Decision Tree C4.5* dengan *Backward Elimination*.

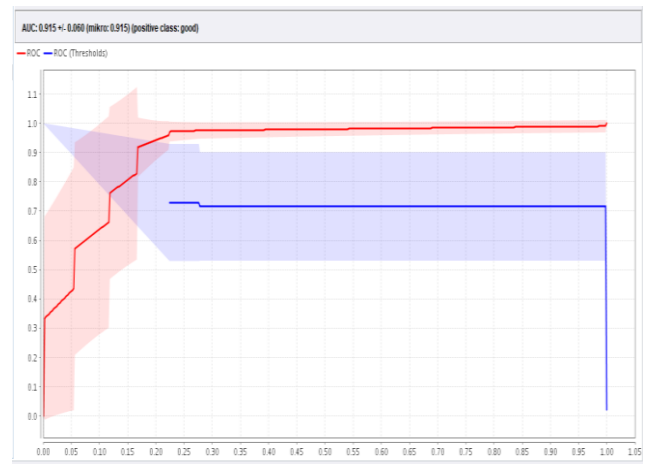
accuracy: 94.00%			
	true bad	true good	class precision
pred. bad	158	4	97.53%
pred. good	21	298	93.42%
class recall	88.27%	98.68%	

Sumber: Hasil Pengolahan Menggunakan RapidMiner (2025)

## 2. Kurva ROC

### a. Algoritma Decision Tree C4.5

Kurva ROC untuk algoritma Decision Tree C4.5 dapat dilihat pada Gambar 2 berikut.



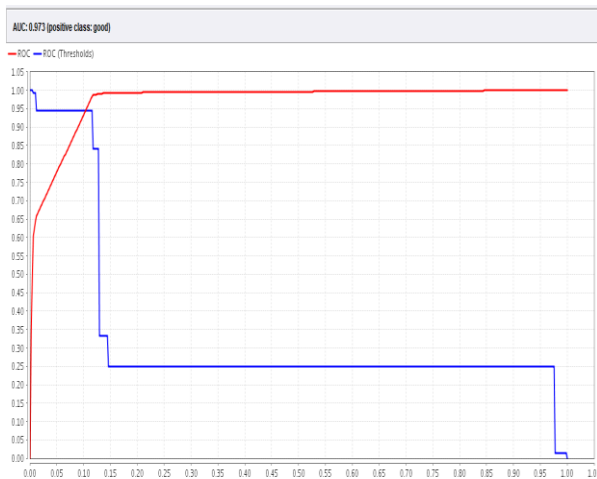
Sumber: Hasil Pengolahan Menggunakan RapidMiner (2025)

Gambar 2 Kurva ROC Algoritma Decision Tree C4.5

Kurva ROC pada Gambar 2 merepresentasikan informasi dari *confusion matrix*. Sumbu horizontal menunjukkan nilai *false positive*, sedangkan sumbu vertikal menggambarkan *true positive*.

### b. Algoritma Decision Tree C4.5 dengan *Backward Elimination*

Kurva ROC untuk algoritma *Decision Tree C4.5* dengan *Backward Elimination* seperti ditunjukkan oleh gambar 3 di bawah ini.



Sumber: Hasil Pengolahan Menggunakan RapidMiner (2025)

Gambar 3 Kurva ROC Metode *Decision Tree C4.5* dengan *Backward Elimination*.

### 3. Analisis Hasil

Perbandingan nilai *accuracy*, *precision*, *recall* dan *AUC* untuk algoritma *Decision Tree C4.5* dan *Decision Tree C4.5* dengan *Backward Elimination* ditunjukkan oleh tabel 4 di bawah ini.

Tabel 4 Evaluasi dan Validasi Model

Algoritma	Accuracy	Precision	Recall	AUC
<i>Decision Tree C4.5</i>	91.90%	91.69%	96.03%	0.915
<i>Decision Tree C4.5</i> dengan <i>Backward Elimination</i>	94.80%	93.42%	98.68%	0.973

Sumber: Hasil Pengolahan Menggunakan RapidMiner (2025)

Tabel 4 membandingkan *accuracy*, *precision*, *recall* dan *AUC* dari algoritma *Decision Tree C4.5* tanpa *backward elimination* dan *Decision Tree C4.5* dengan *backward elimination*. Terlihat bahwa nilai *accuracy* algoritma *Decision Tree C4.5* dengan *backward elimination* memiliki akurasi yang lebih tinggi dibandingkan dengan algoritma *Decision Tree C4.5* tanpa *backward elimination*. Begitupun dengan nilai *AUC*, jelas bahwa nilai *AUC* algoritma *Decision Tree C4.5* dengan *backward elimination* memiliki nilai yang lebih tinggi dibandingkan dengan algoritma *Decision*

*Tree C4.5* tanpa *backward elimination*. Dalam konteks klasifikasi pada data mining, nilai *AUC* dapat dikelompokkan ke dalam beberapa kategori. (Gorunescu, 2011).

- 0.90-1.00 = klasifikasi sangat baik
- 0.80-0.90 = klasifikasi baik
- 0.70-0.80 = klasifikasi cukup
- 0.60-0.70 = klasifikasi buruk
- 0.50-0.60 = klasifikasi salah

Berdasarkan pengelompokan tersebut, dapat disimpulkan bahwa model *Decision Tree C4.5* maupun *Decision Tree C4.5* yang dipadukan dengan *backward elimination* masuk dalam kategori klasifikasi yang sangat baik.

## KESIMPULAN

Berdasarkan hasil penelitian dan pengujian, model *Decision Tree C4.5* tanpa penerapan *backward elimination* menghasilkan akurasi sebesar 91,90% dengan nilai *area under the curve* (*AUC*) sebesar 0,915 dalam menilai kelayakan kredit. Sementara itu, model *Decision Tree C4.5* yang dioptimasi menggunakan *backward elimination* mencapai akurasi 94,80% dengan *AUC* sebesar 0,973. Temuan ini menunjukkan bahwa penggunaan *backward elimination* mampu meningkatkan kinerja metode klasifikasi tersebut.

## REFERENSI

- Agustian, A. A., & Bisri, A. (2019). Data Mining Optimization Using Sample Bootstrapping and Particle Swarm Optimization in the Credit Approval Classification. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(1), 18–27. <https://doi.org/10.24014/ijaidm.v2i1.6299>
- Alfian, A. B., & Nugroho, A. H. D. (2024). Analisis Sistem Pengendalian Internal Terhadap Efektivitas Pemberian Kredit Kendaraan Bermotor di PT. Indomobil Finance Indonesia Cabang Semarang. *Journal of Economic, Bussines and Accounting (COSTING)*, 7(4), 9071–9084. <https://doi.org/10.31539/costing.v7i4.8902>
- Amrin, A.-, & Pahlevi, O.-. (2022). Implementation of Logistic Regression Classification Algorithm and Support Vector Machine for Credit Eligibility Prediction. *Journal of Informatics and Telecommunication Engineering*, 5(2), 433–441. <https://doi.org/10.31289/jite.v5i2.6220>
- Chern, C.-C., Lei, W.-U., Huang, K.-L., & Chen, S.-Y. (2021). A decision tree classifier for credit assessment problems in big data environments. *Information Systems and E-Business Management*, 19(1), 363–386. <https://doi.org/10.1007/s10257-021-00511-w>
- Gorunescu, F. (2011). *Data Mining: Concepts, Models, and Techniques*. Springer.
- Himberg, T. (2020). *Loan Default Prediction with Machine Learning: Vol. 11985 LNAI*. [https://doi.org/10.1007/978-3-030-37720-5\\_4](https://doi.org/10.1007/978-3-030-37720-5_4)
- Jin, Y., Zhang, W., Wu, X., Liu, Y., & Hu, Z. (2021). A Novel Multi-Stage Ensemble Model with a Hybrid Genetic Algorithm for Credit Scoring on Imbalanced Data. *IEEE Access*, 9, 143593–143607. <https://doi.org/10.1109/ACCESS.2021.3120086>
- Lasena, M., & Ahmad, S. R. (2023). Sistem Pendukung Keputusan Kelayakan Pemberian Kredit Nasabah Dengan Metode Electre. *Bulletin of Information Technology (BIT)*, 4(2), 232–238. <https://doi.org/10.47065/bit.v4i2.690>
- Novichasari, S. I. (2021). Peningkatan Akurasi

- Kelayakan Kredit Menggunakan Particle Swarm Optimization. *Jurnal Multimatrix*, 3(1), 86–90. <https://jurnal.unw.ac.id/index.php/mm/article/view/1533%0Ahttps://jurnal.unw.ac.id/index.php/mm/article/view/1533/992>
- Oktafriani, Y., Firmansyah, G., Tjahjono, B., & Widodo, A. M. (2023). Analysis of Data Mining Applications for Determining Credit Eligibility Using Classification Algorithms C4.5, Naïve Bayes, K-NN, and Random Forest. *Asian Journal of Social and Humanities*, 1(12), 1139–1158. <https://doi.org/10.59888/ajosh.v1i12.119>
- Putry, J. B. E., Sasongko, A. T., & Hadikristanto, W. (2024). Optimasi Decision Tree Menggunakan Particle Swarm Optimization (PSO) pada Risiko Kredit KMG Bank DKI. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1403–1410. <https://doi.org/10.57152/malcom.v4i4.1521>
- Religia, Y., Nugroho, A., & Hadikristanto, W. (2021). Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 187–192. <https://doi.org/10.29207/resti.v5i1.2813>
- Song, Y. (2023). Enterprise Credit Rating Prediction Model Based on Data Mining Algorithm. In J. C. Hung, J.-W. Chang, & Y. Pei (Eds.), *Innovative Computing Vol 1 - Emerging Topics in Artificial Intelligence* (pp. 745–751). Springer Nature Singapore.
- Ubaedi, I., & Djaksana, Y. M. (2022). Optimasi Algoritma C4.5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Kredit. *JSiI (Jurnal Sistem Informasi)*, 9(1), 17–26. <https://doi.org/10.30656/jsii.v9i1.3505>
- Ziamba, P., Becker, J., Becker, A., Radomska-Zalas, A., Pawluk, M., & Wierzba, D. (2021). Credit decision support based on real set of cash loans using integrated machine learning algorithms. *Electronics (Switzerland)*, 10(17), 1–22. <https://doi.org/10.3390/electronics10172099>