

PENERAPAN ALGORITMA DECISION TREE UNTUK KLASIFIKASI PERSETUJUAN PINJAMAN SEBAGAI PENDUKUNG PENGAMBILAN KEPUTUSAN RISIKO KREDIT

Moammar Iqbal¹, Irfan Ali², Bani Nurhakim³, Puji Pramudya Marta⁴.

Program Studi Komputerisasi Akuntansi¹
Program Studi Rekayasa Perangkat Lunak²
Program Studi Manajemen Informatika³
Program Studi Sistem Informasi⁴

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
ikbal30042005@gmail.com

(*) Corresponding Author : ikbal30042005@gmail.com.
Published : 30 Maret 2026

Abstract—Competition in lending compels financial institutions to conduct credit assessments that are fast, objective, and accountable in order to minimize credit risk. One of the recurring root problems is the reliance on subjective judgment, which makes it difficult to maintain consistency and transparency in decision-making. This study examines the application of a classification model to support loan approval decisions with the aim of improving the objectivity and consistency of credit risk assessment. The identified root cause is the dependence on manual assessment practices that are prone to subjectivity, thereby increasing the risk of non-performing loans; therefore, this research utilizes historical data from Kaggle (a sample of 10,000 records) and the CRISP-DM framework to develop an interpretable Decision Tree model. The process includes data type adjustment, identifier generation, attribute role assignment, parameter optimization using grid search, and evaluation through cross-validation. The results show that the optimized Decision Tree model achieves an accuracy of 90.93%; precision and recall for the “approved” class are 87.70% and 69.67%, respectively, while for the “not approved” class the precision is 91.64% and the recall is 97.14%. The dominant contributing attributes include default history, loan-to-income ratio, interest rate, home ownership, and income. The model produces practical if-then rules for auto-decline policies and further review, thereby offering a balance between predictive performance and interpretability; however, attention to class imbalance is required prior to operational deployment.

Keywords : Loan Approval Classification, Decision Tree, CRISP-DM, Parameter Optimization, Decision Making.

Abstrak—Persaingan dalam pemberian kredit menuntut lembaga keuangan melakukan penilaian kredit yang cepat, objektif, dan dapat dipertanggungjawabkan untuk meminimalkan risiko kredit. Salah satu akar masalah yang sering ditemui adalah ketergantungan pada penilaian subjektif sehingga konsistensi dan transparansi keputusan sulit dijaga. Penelitian ini membahas penerapan model klasifikasi untuk mendukung pengambilan keputusan persetujuan pinjaman dengan tujuan meningkatkan objektivitas dan konsistensi penilaian risiko kredit. Akar permasalahan yang diidentifikasi adalah ketergantungan praktik penilaian manual yang rentan terhadap subjektivitas sehingga meningkatkan risiko kredit bermasalah; oleh karena itu penelitian ini menggunakan data historis dari Kaggle (sampel 10.000 entri) dan kerangka kerja CRISP-DM untuk membangun model *Decision Tree* yang dapat dijelaskan. Proses meliputi penyesuaian tipe data, pembuatan identifier, penetapan peran atribut, optimasi parameter melalui grid search, serta evaluasi dengan *cross-validation*. Hasil menunjukkan model *Decision Tree* teroptimasi mencapai akurasi 90,93%; *precision* dan *recall* untuk kelas “disetujui” masing-masing 87,70% dan 69,67%, sedangkan untuk kelas “tidak disetujui” *precision* 91,64% dan *recall* 97,14%. Atribut yang dominan berkontribusi meliputi riwayat gagal bayar, rasio pinjaman terhadap pendapatan, suku bunga, kepemilikan rumah, dan pendapatan. Model ini menghasilkan aturan *if-then* yang praktikabel untuk kebijakan *auto-decline* dan pemeriksaan lanjutan, sehingga menawarkan keseimbangan antara kinerja prediktif dan interpretabilitas; namun diperlukan perhatian terhadap ketidakseimbangan kelas sebelum penerapan operasional.

Kata Kunci : Klasifikasi Persetujuan Pinjaman, *Decision Tree*, *CRISP-DM*, Optimasi Parameter, Pengambilan Keputusan.

INTRODUCTION

Perkembangan teknologi informasi dan analisis data telah membawa perubahan signifikan dalam proses pengambilan keputusan di sektor keuangan, khususnya dalam kegiatan pemberian kredit. Lembaga keuangan saat ini dihadapkan pada kebutuhan untuk memproses data peminjam dalam jumlah besar secara cepat dan akurat guna meminimalkan risiko kredit serta meningkatkan efisiensi operasional. Pendekatan tradisional yang mengandalkan penilaian manual dan pertimbangan subjektif dinilai kurang efektif dalam menghadapi kompleksitas karakteristik peminjam, sehingga mendorong pemanfaatan teknik machine learning sebagai pendukung pengambilan keputusan kredit berbasis data [1].

Dalam konteks manajemen risiko kredit, keputusan persetujuan pinjaman merupakan tahapan kritis karena kesalahan keputusan dapat berdampak langsung pada peningkatan kredit bermasalah dan kerugian finansial. Oleh karena itu, diperlukan suatu model klasifikasi yang mampu mempelajari pola hubungan antara karakteristik peminjam dan status pinjaman, sehingga dapat memberikan rekomendasi keputusan yang lebih objektif dan konsisten. Berbagai penelitian menunjukkan bahwa algoritma machine learning mampu meningkatkan kualitas prediksi risiko kredit dibandingkan metode konvensional, terutama ketika digunakan pada data historis peminjam yang terstruktur [2].

Penelitian ini menggunakan dataset kredit yang bersumber dari *Kaggle* dan diproses menggunakan perangkat lunak *RapidMiner*. Dataset tersebut terdiri dari 14 atribut dan satu label target, yaitu *loan_status*, yang merepresentasikan status persetujuan pinjaman. Atribut yang digunakan mencerminkan karakteristik demografis dan finansial peminjam, seperti usia, tingkat pendidikan, pendapatan, jumlah pinjaman, tujuan pinjaman, suku bunga, skor kredit, serta riwayat kredit sebelumnya. Ringkasan atribut yang digunakan dalam penelitian ini disajikan pada **Error! Reference source not found.**, yang memberikan gambaran struktur data serta peran masing-masing variabel dalam proses pemodelan.

Meskipun banyak algoritma *machine learning* mampu menghasilkan performa prediksi yang tinggi, penerapan model dalam lingkungan operasional lembaga keuangan tidak

hanya menuntut akurasi, tetapi juga menuntut kemudahan interpretasi hasil model. Keputusan kredit harus dapat dijelaskan secara logis kepada petugas kredit dan pihak manajemen, serta dapat dijadikan dasar pertanggungjawaban dalam proses audit dan evaluasi kebijakan. Oleh karena itu, model yang bersifat *black-box* sering kali sulit diimplementasikan secara langsung dalam praktik, meskipun memiliki performa statistik yang baik [3].

Akar permasalahan yang melatarbelakangi penelitian ini adalah belum optimalnya pemanfaatan model klasifikasi yang menghasilkan aturan keputusan yang jelas dan mudah dipahami dalam proses persetujuan pinjaman. Tanpa adanya aturan keputusan yang terstruktur, proses penilaian kredit berpotensi menjadi tidak konsisten dan sangat bergantung pada penilaian individu. Kondisi tersebut dapat menyebabkan perbedaan keputusan pada kasus yang serupa serta memperlambat proses persetujuan pinjaman. Oleh karena itu, diperlukan pendekatan yang tidak hanya mampu melakukan klasifikasi secara akurat, tetapi juga mampu menghasilkan aturan keputusan yang dapat diterapkan secara operasional. [4], [5], [6]

Berdasarkan permasalahan tersebut, penelitian ini memilih algoritma *Decision Tree* sebagai model klasifikasi utama. Algoritma *Decision Tree* memiliki keunggulan dalam menghasilkan struktur pohon keputusan yang dapat diterjemahkan ke dalam aturan *if-then* yang sederhana dan mudah dipahami oleh pengguna non-teknis. Selain itu, *Decision Tree* banyak digunakan dalam penelitian kredit karena kemampuannya menjelaskan kontribusi setiap atribut terhadap keputusan akhir, sehingga sesuai dengan kebutuhan interpretabilitas dalam manajemen risiko kredit [1], [7].

Kebaharuan penelitian ini terletak pada penerapan algoritma *Decision Tree* yang diimplementasikan secara sistematis menggunakan *RapidMiner*, disertai dokumentasi alur pemodelan dan pengaturan parameter secara rinci, serta interpretasi hasil model ke dalam bentuk aturan keputusan yang dapat digunakan sebagai pendukung pengambilan keputusan risiko kredit. Dengan pendekatan tersebut, penelitian ini tidak hanya berfokus pada aspek teknis pemodelan, tetapi juga pada aspek praktis penerapan model dalam proses persetujuan pinjaman [8], [9], [10].

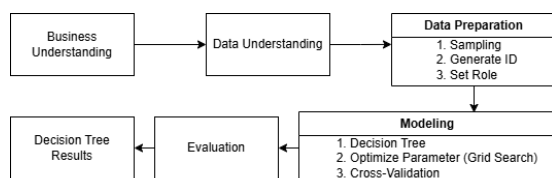
Dengan demikian, tujuan utama penelitian ini adalah membangun model klasifikasi berbasis *Decision Tree* untuk memprediksi status persetujuan pinjaman menggunakan data historis kredit, serta menyediakan dasar pengambilan keputusan yang lebih objektif, konsisten, dan

mudah dipahami bagi pihak yang terlibat dalam manajemen risiko kredit.

MATERIALS AND METHODS

Sumber data penelitian ini adalah dataset publik berjudul Loan Approval Classification Data yang tersedia pada platform Kaggle. Dataset tersebut disediakan oleh kontributor pada laman Kaggle dan dapat diakses secara bebas untuk tujuan penelitian akademik. Saya memilih dataset ini karena memenuhi kriteria penelitian: berisi atribut demografis, finansial, dan riwayat kredit yang relevan dengan permasalahan klasifikasi persetujuan pinjaman (`loan_status`), terdokumentasi dalam format terstruktur (CSV), serta sudah umum digunakan pada studi-studi sejenis sehingga memudahkan replikasi hasil penelitian.

Tahapan perancangan penelitian ini saya susun berdasarkan adaptasi metodologi *CRISP-DM (Cross Industry Standard Process for Data Mining)* yang disesuaikan dengan kebutuhan eksperimen klasifikasi persetujuan pinjaman menggunakan algoritma *Decision Tree*. *CRISP-DM* merupakan kerangka kerja yang telah banyak digunakan dan diakui sebagai metodologi standar dalam proyek data mining dan ilmu data karena kemampuannya menyediakan struktur yang sistematis dari awal hingga akhir proyek, termasuk fase pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi hasil [11]. Secara garis besar, alur perancangan yang saya terapkan mengikuti urutan seperti pada Gambar 1. Setiap tahapan dirancang agar dapat direplikasi dan terdokumentasi dengan baik sesuai bukti artefak eksperimen (dataset, konfigurasi operator, parameter grid, dan hasil performa) yang tercatat selama pelaksanaan penelitian. Rincian tahapan perancangan adalah sebagai berikut.



Gambar 1. Alur Penelitian

Tahap Business Understanding bertujuan untuk memahami tujuan utama penelitian yang berkaitan dengan pemanfaatan data dalam mendukung pengambilan keputusan bisnis. Dalam penelitian ini, fokus utama adalah membangun model klasifikasi yang mampu memprediksi status persetujuan pinjaman berdasarkan karakteristik peminjam. Dengan

memahami kebutuhan bisnis tersebut, proses analisis data dapat diarahkan untuk menghasilkan model yang relevan dan bermanfaat dalam pengelolaan risiko kredit. Hasil dari tahap ini menjadi dasar dalam menentukan pendekatan analisis, pemilihan algoritma, serta strategi pemodelan yang akan digunakan. Dengan demikian, model yang dihasilkan diharapkan mampu membantu lembaga keuangan dalam menilai kelayakan kredit secara lebih objektif, konsisten, dan berbasis data.

Tahap Data Understanding bertujuan untuk memperoleh pemahaman menyeluruh mengenai dataset yang digunakan dalam penelitian. Pada tahap ini dilakukan pengamatan terhadap struktur data, jenis atribut, serta kualitas data yang tersedia. Proses ini penting untuk memastikan bahwa dataset sesuai dengan kebutuhan penelitian dan dapat digunakan dalam proses pemodelan klasifikasi. Selain itu, tahap ini juga membantu mengidentifikasi potensi permasalahan data, seperti nilai yang tidak konsisten, data kosong, maupun distribusi nilai atribut. Dengan memahami karakteristik dataset sejak awal, peneliti dapat merencanakan proses pra-pemrosesan data secara tepat. Hasil dari tahap ini menjadi dasar dalam menentukan langkah-langkah yang diperlukan pada fase Data Preparation sebelum proses pemodelan dilakukan.

Tahap Data Preparation bertujuan untuk menyiapkan dataset agar siap digunakan dalam proses pemodelan klasifikasi. Pada tahap ini dilakukan beberapa proses penting, seperti penyesuaian tipe data pada setiap atribut agar sesuai dengan karakteristik nilainya, baik numerik maupun kategorikal. Selanjutnya dilakukan proses sampling dengan mengambil 10.000 data secara acak dari dataset awal untuk meningkatkan efisiensi komputasi selama proses pemodelan. Setelah itu dibuat ID nominal sebagai identitas unik setiap record untuk memudahkan pelacakan data. Peneliti juga menetapkan peran atribut dengan menentukan `loan_status` sebagai label dan atribut lainnya sebagai fitur. Selain itu dilakukan pemeriksaan missing value untuk memastikan dataset tidak mengandung data kosong sehingga siap digunakan dalam proses modeling.

Tahap Modeling merupakan proses pembangunan model klasifikasi menggunakan algoritma Decision Tree. Dataset yang telah melalui tahap Data Preparation digunakan sebagai data masukan dalam proses pembelajaran model. Variabel `loan_status` ditetapkan sebagai label yang akan diprediksi, sedangkan atribut lainnya digunakan sebagai variabel prediktor. Untuk memperoleh kinerja model yang optimal, dilakukan proses optimasi parameter pada algoritma Decision

Tree dengan menguji berbagai kombinasi parameter yang mempengaruhi struktur pohon keputusan. Proses ini dikombinasikan dengan teknik cross-validation untuk mengevaluasi performa model secara lebih objektif. Hasil dari proses ini adalah model Decision Tree terbaik yang memiliki keseimbangan antara akurasi, kompleksitas model, dan kemampuan generalisasi terhadap data baru.

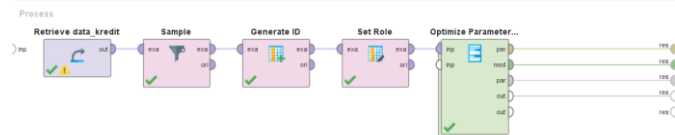
Tahap Evaluation bertujuan untuk menilai kinerja model klasifikasi yang telah dibangun pada tahap sebelumnya. Evaluasi dilakukan menggunakan beberapa metrik performa klasifikasi, yaitu accuracy, precision, dan recall, untuk memberikan gambaran menyeluruh mengenai kualitas model. Selain itu digunakan confusion matrix untuk melihat perbandingan antara hasil prediksi model dengan data aktual. Melalui evaluasi ini dapat diketahui sejauh mana model Decision Tree mampu mengklasifikasikan status persetujuan pinjaman secara tepat. Hasil evaluasi kemudian dibandingkan dengan tujuan penelitian yang telah ditetapkan. Apabila performa model telah memenuhi kriteria yang diharapkan, maka model dianggap layak digunakan sebagai dasar analisis dan interpretasi hasil penelitian.

Tahap Decision Tree Results merupakan tahap akhir yang berfokus pada penyajian dan dokumentasi keluaran model Decision Tree terbaik. Pada tahap ini dilakukan pengamatan terhadap struktur pohon keputusan yang dihasilkan, seperti jumlah node, kedalaman pohon, serta atribut yang digunakan dalam proses pemisahan. Selain itu, aturan keputusan yang terbentuk dari jalur pohon juga dianalisis dalam bentuk aturan if-then yang menunjukkan kondisi tertentu yang mengarah pada keputusan persetujuan atau penolakan pinjaman. Peneliti juga mengidentifikasi atribut yang paling sering muncul dalam pemisahan node sebagai indikasi faktor penting dalam penilaian kredit. Seluruh hasil tersebut didokumentasikan sebagai dasar untuk proses interpretasi dan pembahasan pada bab selanjutnya.

RESULTS AND DISCUSSION

Tahap *Modeling* difokuskan pada prosedur pencarian konfigurasi parameter terbaik untuk algoritma *Decision Tree* melalui pendekatan optimasi terstruktur. Untuk tujuan ini digunakan operator *Optimize Parameters (Grid)* pada *RapidMiner* yang menjalankan eksplorasi kombinasi parameter secara sistematis.

Operator tersebut dibungkus oleh skema *Cross Validation* sehingga setiap kombinasi parameter dinilai secara internal melalui pembagian data menjadi lipatan-lipatan (*folds*), sehingga pemilihan parameter dilakukan berdasarkan kinerja rata-rata pada proses validasi silang. Diagram proses yang menunjukkan susunan operator dari *Retrieve* → *Sample* → *Generate ID* → *Set Role* → *Optimize Parameters* disertakan untuk dokumentasi proses eksperimen.

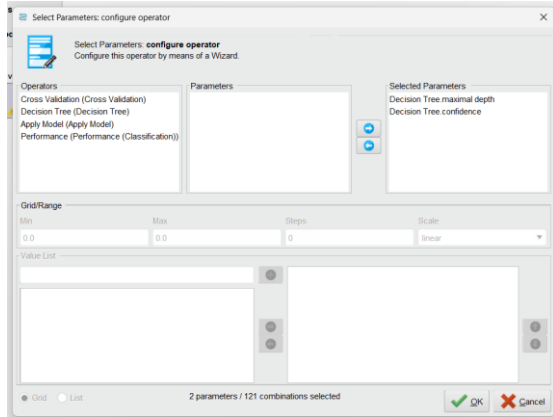


Gambar 2 Proses *Optimize Parameters* pada canvas *RapidMiner*. Sumber: Diolah penulis menggunakan RapidMiner

Optimize Parameters (Grid) dikonfigurasi untuk mengeksplorasi dua parameter utama *Decision Tree* yang dipilih untuk dioptimalkan berdasarkan literatur dan sifat algoritma: *maximal_depth* (kedalaman maksimum pohon) dan *confidence* (tingkat kepercayaan untuk mekanisme pruning). Pemilihan parameter ini didasarkan pada prinsip bahwa *maximal_depth* mengendalikan kompleksitas struktur pohon sehingga berkaitan langsung dengan risiko *underfitting/overfitting*, sedangkan *confidence* memengaruhi proses pruning yang mengurangi kompleksitas model pasca-pembentukan pohon untuk meningkatkan generalisasi. Oleh karena itu, fokus tuning diarahkan pada parameter tersebut untuk menemukan titik keseimbangan antara kompleksitas dan kemampuan generalisasi pohon keputusan.

Untuk *Decision Tree.maximal_depth* digunakan rentang pencarian dari *min* = -1 hingga *max* = 100 dengan *steps* = 10 pada konfigurasi grid. Nilai -1 ditetapkan sebagai tanda tidak ada batasan kedalaman (*unlimited*) pada beberapa implementasi, sedangkan nilai positif pada grid merepresentasikan batasan kedalaman yang bertingkat meningkat; penggunaan langkah (*steps*) menghasilkan sejumlah nilai diskret yang akan diuji sepanjang rentang tersebut. Untuk *Decision Tree.confidence* rentang yang diuji adalah *min* = 1.0E-7 hingga *max* = 0.5 dengan *steps* = 10, sehingga menghasilkan nilai-nilai *confidence* yang terdistribusi secara linear dari sangat kecil hingga 0.5. Kombinasi dari kedua parameter ini membentuk cartesian product grid sehingga operator *Optimize Parameters* menghasilkan sejumlah kombinasi parameter yang dievaluasi (konfigurasi saat ini menghasilkan 121 kombinasi parameter yang diuji secara berulang melalui mekanisme *cross-validation*). Konfigurasi grid

tersebut dipilih untuk memberi cakupan yang memadai pada ruang parameter sambil mempertahankan jumlah percobaan yang masih dapat ditangani dalam kerangka komputasi yang tersedia.



Gambar 3 Konfigurasi grid parameter yang diuji
Sumber: Diolah penulis menggunakan RapidMiner

Di dalam operator *Optimize Parameters*, *Cross Validation* bertindak sebagai *wrapper* yang memecah data sampel ke dalam beberapa *fold* (stratifikasi dipakai bila diperlukan untuk menjaga proporsi kelas) sehingga setiap kombinasi parameter dilatih dan divalidasi bergantian pada subset data yang berbeda. Secara praktis, skema ini terdiri dari blok pelatihan (di mana operator *Decision Tree* dijalankan dengan satu kombinasi parameter) dan blok pengujian (di mana *operator Apply Model* dan *Performance* menghitung metrik untuk *fold* tersebut). Seperti yang dapat dilihat pada struktur ini direpresentasikan secara visual pada canvas RapidMiner dan pada tingkat operator internal menampilkan aliran *Decision Tree* → *Apply Model* → *Performance* untuk tiap *fold*. Pendekatan ini memastikan bahwa pemilihan parameter tidak bergantung pada satu pembagian data saja tetapi dilandasi hasil agregat dari beberapa pembagian yang berbeda.

Dengan menggunakan konfigurasi parameter tersebut, model *Decision Tree* menghasilkan nilai akurasi sebesar 90,93% dengan deviasi sebesar ±1,20% berdasarkan rata-rata mikro (*micro average*) dari proses *cross validation*. Nilai akurasi ini menunjukkan bahwa model mampu mengklasifikasikan status persetujuan pinjaman dengan tingkat ketepatan yang tinggi pada berbagai subset data uji selama proses validasi silang. Hasil performa model beserta konfigurasi parameter terbaik dapat diamati secara rinci pada keluaran *ParameterSet* yang ditampilkan pada Gambar .

ParameterSet

```
Parameter set:

Performance:
PerformanceVector [
----accuracy: 90.93% +/- 1.20% (micro average: 90.93%)
ConfusionMatrix:
True: 1 0
1: 1576 221
0: 686 7517
]
Decision Tree.maximal_depth = 19
Decision Tree.confidence = 0.35000002999999996
```

Gambar 4 Hasil ParameterSet dari proses optimasi parameter *Decision Tree*

Dengan demikian, parameter *maximal_depth* = 19 dan *confidence* = 0,35 ditetapkan sebagai konfigurasi terbaik.

Evaluation

Tahap evaluasi dilakukan untuk menilai kinerja model klasifikasi *Decision Tree* yang telah dibangun menggunakan parameter terbaik hasil proses optimasi. Evaluasi model bertujuan untuk mengukur sejauh mana model mampu melakukan klasifikasi status persetujuan pinjaman secara tepat berdasarkan data uji, dengan menggunakan metrik evaluasi yang relevan, yaitu akurasi, *precision*, dan *recall*. Seluruh hasil evaluasi diperoleh dari proses *cross validation* dan ditampilkan dalam bentuk *confusion matrix* serta ringkasan metrik performa.

Berdasarkan hasil evaluasi, model *Decision Tree* menghasilkan akurasi sebesar 90,93% dengan deviasi sebesar ±1,20% berdasarkan *micro average*. Nilai akurasi ini menunjukkan bahwa secara keseluruhan model mampu mengklasifikasikan status persetujuan pinjaman dengan tingkat ketepatan yang tinggi. Ringkasan nilai akurasi, *confusion matrix*, serta metrik *precision* dan *recall* untuk masing-masing kelas ditampilkan secara visual pada Gambar .

accuracy: 90.93% +/- 1.20% (micro average: 90.93%)

	true 1	true 0	cl
pred. 1	1576	221	8
pred. 0	686	7517	9
class recall	69.67%	97.14%	

Gambar 5 Hasil evaluasi model *Decision Tree*

Analisis *confusion matrix* menunjukkan distribusi hasil prediksi terhadap kondisi aktual sebagai berikut: sebanyak 1.576 data diklasifikasikan dengan benar sebagai kelas 1 (pinjaman disetujui), dan 7.517 data diklasifikasikan dengan benar sebagai kelas 0 (pinjaman tidak disetujui). Sementara itu, terdapat 221 data yang diprediksi sebagai disetujui tetapi

sebenarnya tidak disetujui (*false positive*), serta 686 data yang diprediksi tidak disetujui padahal sebenarnya disetujui (*false negative*). Distribusi ini menggambarkan pola kesalahan klasifikasi yang terjadi pada kedua kelas.

Berdasarkan *confusion matrix* tersebut, diperoleh nilai *precision* sebesar 87,70% untuk kelas 1 dan 91,64% untuk kelas 0. Nilai *precision* yang lebih tinggi pada kelas 0 (pinjaman tidak disetujui) menunjukkan bahwa sebagian besar prediksi “tidak disetujui” yang dihasilkan oleh model memang sesuai dengan kondisi aktual. Sementara itu, nilai *precision* pada kelas 1 (pinjaman disetujui) menunjukkan bahwa mayoritas prediksi “disetujui” juga berada pada kelas yang benar, meskipun masih terdapat sejumlah kesalahan prediksi.

Nilai *recall* untuk kelas 1 sebesar 69,67%, sedangkan *recall* untuk kelas 0 mencapai 97,14%. Perbedaan yang cukup signifikan antara nilai *recall* kedua kelas ini menunjukkan bahwa model lebih sensitif dalam mengenali pemohon pinjaman yang termasuk dalam kategori tidak disetujui dibandingkan dengan pemohon yang disetujui. Artinya, sebagian besar kasus aktual kelas 0 berhasil diidentifikasi dengan baik oleh model, sedangkan pada kelas 1 masih terdapat sejumlah kasus yang tidak tertangkap oleh model.

Ketidakseimbangan nilai *recall* ini dapat dipahami baik dari sisi karakteristik data maupun dari sudut pandang bisnis. Secara data, distribusi kelas menunjukkan bahwa kelas tidak disetujui memiliki jumlah data yang lebih besar, sehingga model cenderung lebih optimal dalam mempelajari pola pada kelas tersebut. Dari perspektif manajemen risiko kredit, sensitivitas yang tinggi terhadap kelas tidak disetujui dapat dianggap sebagai karakteristik yang diinginkan, karena membantu meminimalkan risiko pemberian pinjaman kepada peminjam yang berpotensi gagal bayar. Dengan kata lain, model bersifat lebih konservatif dalam memberikan persetujuan pinjaman, meskipun konsekuensinya adalah berkurangnya kemampuan model dalam menangkap seluruh pemohon yang sebenarnya layak disetujui.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model *Decision Tree* memiliki kinerja yang baik dengan tingkat akurasi yang tinggi serta kemampuan yang sangat kuat dalam mengidentifikasi pemohon berisiko. Interpretasi menyeluruh terhadap nilai akurasi, *precision*, dan *recall*, serta pola kesalahan klasifikasi yang ditunjukkan oleh *confusion matrix*, memberikan gambaran yang

jelas mengenai karakteristik performa model dalam konteks penilaian persetujuan pinjaman.

CONCLUSION

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Penelitian ini berhasil menerapkan algoritma *Decision Tree* untuk melakukan klasifikasi persetujuan pinjaman berdasarkan data historis kredit yang bersumber dari *Kaggle*. Proses klasifikasi dilakukan dengan memanfaatkan sejumlah atribut demografis dan finansial peminjam, seperti usia, pendapatan, kepemilikan tempat tinggal, persentase pinjaman terhadap pendapatan, suku bunga pinjaman, skor kredit, serta riwayat gagal bayar sebelumnya, dengan variabel *loan_status* sebagai label kelas yang merepresentasikan status persetujuan pinjaman.
2. Proses pemodelan yang dilengkapi dengan optimasi parameter menggunakan *grid search* dan validasi silang (*cross-validation*) menghasilkan kinerja model yang baik, dengan nilai akurasi mikro rata-rata sebesar 90,93%. Parameter terbaik yang diperoleh pada proses optimasi adalah *maximal depth* = 19 dan *confidence* = 0,35, yang menunjukkan bahwa pengaturan parameter yang tepat berperan penting dalam meningkatkan performa klasifikasi model *Decision Tree*.
3. Hasil evaluasi kinerja model menunjukkan bahwa *Decision Tree* memiliki kemampuan yang sangat baik dalam mengidentifikasi peminjam yang berisiko atau tidak disetujui, ditunjukkan oleh nilai *recall* yang tinggi pada kelas penolakan pinjaman. Kondisi ini selaras dengan tujuan manajemen risiko kredit, yaitu meminimalkan potensi gagal bayar. Selain itu, struktur pohon keputusan yang dihasilkan mampu memberikan aturan keputusan yang jelas dan mudah dipahami, sehingga model yang dibangun tidak hanya memiliki performa yang baik secara statistik, tetapi juga dapat digunakan sebagai pendukung pengambilan keputusan risiko kredit yang bersifat objektif, konsisten, dan dapat dipertanggungjawabkan.

REFERENCE

- [1] V. Chang, S. Sivakulasingam, H. Wang, S. T. Wong, M. A. Ganatra, and J. Luo, “Credit risk prediction using machine learning and deep learning: A study on credit card customers,” *Risks*, vol. 12, no. 11, p. 174, 2024, doi: 10.3390/risks12110174.
- [2] E. Dumitrescu, S. Hué, C. Hurlin, and S. Tokpavi, “Machine learning for credit scoring: Improving logistic regression with

- non-linear decision-tree effects,” *European Journal of Operational Research*, vol. 297, no. 3, pp. 1178–1192, 2022, doi: 10.1016/j.ejor.2021.06.053.
- [3] Y. Chen, R. Calabrese, and B. Martín-Barragán, “Interpretable machine learning for credit scoring datasets,” *European Journal of Operational Research*, vol. 312, no. 2, pp. 357–372, 2024, doi: 10.1016/j.ejor.2023.06.036.
- [4] I. D. Mienye, Y. Sun, and Z. Wang, “Prediction performance of improved decision tree-based algorithms: A review,” *Procedia Manuf.*, vol. 35, pp. 698–703, 2019, doi: 10.1016/j.promfg.2019.06.011.
- [5] J. R. Quinlan, “Induction of decision trees,” *Mach. Learn.*, vol. 1, pp. 81–106, 1986.
- [6] Nadia Tiara Rahman, “ANALISA ALGORITMA DECISION TREEDAN NAÏVE BAYESPADA PASIEN PENYAKIT LIVER,” *Jurnal Fasilkom*, 2020.
- [7] C. C. Chern, “A decision tree classifier for credit assessment problems in big data environments,” *Information Systems Frontiers*, 2021, doi: 10.1007/s10257-021-00511-w.
- [8] D. Haganta Depari *et al.*, “Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung,” *JURNAL INFORMATIK Edisi ke*, vol. 18, p. 2022.
- [9] I. Setiawati, A. P. Wibowo, A. Hermawan, M. Teknologi, I. Universitas, and T. Yogyakarta, “IMPLEMENTASI DECISION TREE UNTUK MENDIAGNOSIS PENYAKIT LIVER,” 2019.
- [10] I. Setiawati, A. P. Wibowo, A. Hermawan, M. Teknologi, I. Universitas, and T. Yogyakarta, “IMPLEMENTASI DECISION TREE UNTUK MENDIAGNOSIS PENYAKIT LIVER,” 2019.
- [11] H. Acuña-Cid, “Integration of the CRISP-DM Model with Network Analysis,” *MDPI*, 2025, doi: 10.3390/someDOI.