

Prediction of Customer Creditworthiness with the C4.5 Algorithm at PT Menara Indonesia Company

¹ Dhoni Hanif Supriyadi, ² Fernando B Siahaan, ³ Sriyadi, ⁴ Syaiful Anwar, ⁵ Felix Wuryo Handono

^{1,2,3,4,5}Universitas Bina Sarana Informatika, Indonesia

[1dhonihanif354@gmail.com](mailto:dhonihanif354@gmail.com); [2fernando.fbs@bsi.ac.id](mailto:fernando.fbs@bsi.ac.id); [3sriyadi.sry@bsi.ac.id](mailto:sriyadi.sry@bsi.ac.id); [4syaiful.sfa@bsi.ac.id](mailto:syaiful.sfa@bsi.ac.id); [5felix@bsi.ac.id](mailto:felix@bsi.ac.id)

Article Info

Article history:

Received, xxx xx xxxx

Revised, xxx xx xxxx

Accepted, xxx xx xxx

Kata Kunci:

Prediksi

Kredit Nasabah

Algoritma C.4.5

Keywords:

Prediction

Customer Credit

C.4.5 Algorithm

ABSTRAK

Penilaian kredit nasabah masih dilakukan dengan metode tradisional yang memakan waktu lama dan kurang akurat. Hal ini terbukti dengan masih adanya nasabah dengan kredit bermasalah yang lolos dalam proses pengajuan pinjaman. Untuk mengatasi masalah ini, penelitian ini bertujuan untuk mengembangkan metode menggunakan algoritma C4.5. Tujuan dari penelitian ini adalah untuk meningkatkan efisiensi dan efektivitas dalam proses penagihan hutang perusahaan. Penelitian ini menggunakan data nasabah dari PT Menara Indonesia yang memiliki pinjaman kredit. Data tersebut mencakup tujuh variabel independen: pendapatan bersih, jumlah pinjaman, skor kredit, jumlah tunggakan, tenor, aset, dan usia pinjaman, serta satu variabel dependen yaitu risiko kredit. Algoritma C4.5 diterapkan untuk membangun model prediksi pelunasan kredit nasabah. Model ini diuji menggunakan metode k-fold cross validation dengan nilai k = 10. Hasil pengujian menunjukkan bahwa model C4.5 memiliki performa yang sangat baik, dengan akurasi sebesar 99,70%, presisi 99,25%, recall 98,52%, dan f1-score 98,88%. Kelebihan dari metode ini adalah kemampuannya untuk memberikan prediksi yang sangat akurat, sehingga dapat membantu perusahaan dalam mengidentifikasi nasabah yang berisiko tinggi dan meningkatkan proses penagihan hutang secara keseluruhan.

ABSTRACT

Customer credit assessment is still carried out using traditional methods that are time-consuming and less accurate. This is evidenced by the fact that there are still customers with problematic credit who pass the loan application process. To address this issue, this research aims to develop a method using the C4.5 algorithm. The purpose of this research is to improve the efficiency and effectiveness of the company's debt collection process. This research uses customer data from PT Menara Indonesia, which has credit loans. The data includes seven independent variables: net income, loan amount, credit score, number of arrears, tenure, assets, and loan age, as well as one dependent variable, namely credit risk. The C4.5 algorithm is applied to build a customer credit repayment prediction model. This model is tested using the k-fold cross-validation method with k = 10. The test results show that the C4.5 model has excellent performance, with an accuracy of 99.70%, precision of 99.25%, recall of 98.52%, and an F1-score of 98.88%. The advantage of this method is its ability to provide highly accurate predictions, thereby helping the company identify high-risk customers and improve the overall debt collection process.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



Penulis Korespondensi:

Fernando B Siahaan,

Program Studi Sistem Informatika,

Universitas Bina Sarana Informatika,

Email: fernando.fbs@bsi.ac.id

1. PENDAHULUAN

Salah satu instrumen utama yang digunakan untuk mendukung berbagai kegiatan ekonomi adalah kredit. Kredit adalah fasilitas yang sering dimanfaatkan oleh individu dan perusahaan untuk memenuhi kebutuhan finansial mereka. Risiko kredit mengacu pada potensi kerugian finansial yang dihadapi oleh pemberi pinjaman

ketika pelanggan tidak dapat memenuhi kewajibannya untuk membayar kembali pinjaman beserta bunganya[1].

Pemanfaatan data mining untuk mengklasifikasikan data berguna untuk memprediksi calon nasabah yang akan menunggak atau lincaran dalam pembayaran kredit sesuai dengan masa cicilan yang dipilih. Perusahaan perlu mengklasifikasikan data nasabah mana yang berpotensi untuk di setujui dalam ajun kreditnya [2]. Untuk mengatasi ketergantungan metode tradisional, data mining menawarkan solusi yang lebih akurat dan efisien. Algoritma C4.5 adalah evolusi dari algoritma ID3, yang dapat digunakan untuk mengklasifikasikan kelas dan menggunakan pohon keputusan untuk membuat keputusan. Dalam hal ini, algoritma C4.5 diterapkan untuk menyelesaikan tantangan operasional terkait penilaian kredit. Outputnya adalah pohon keputusan yang menggabungkan data ke dalam sistem atau struktur yang memungkinkan pengambilan keputusan [3].

PT Menara Indonesia dapat melakukan analisis kredit menggunakan C4.5 karena penentuan kelayakan pemberian kredit menggunakan 5C menunjukkan hasil yang tidak akurat dan proses yang lama. Penelitian ini bertujuan untuk mengklasifikasikan antara pemohon kredit yang disetujui dan tidak disetujui, dengan fokus pada peningkatan kecepatan dan akurasi dalam proses penilaian kelayakan kredit.

Beberapa penelitian telah berfokus pada risiko kredit dan menggunakan algoritma C4.5 untuk analisis. Handayani et al. (2021) mengembangkan algoritma C4.5 dan mencapai akurasi sebesar 79% dalam memprediksi risiko kredit menggunakan teknik data mining [4]. meningkatkan evaluasi ini, mencapai tingkat akurasi 71,60% dengan memanfaatkan data kredit dari Jerman menggunakan algoritma C4.5 untuk menentukan kelayakan kredit dan mencapai akurasi 93% dengan nilai AUC klasifikasi yang baik sebesar 0,898 [5]. menerapkan data mining untuk memprediksi transaksi bermasalah dan mencapai tingkat akurasi 65% [2]. Hidayatulloh et al. (2022) menggunakan algoritma C4.5 untuk memprediksi kegagalan pembayaran kredit dan memperoleh tingkat akurasi 93,75% dan 96,77% dengan rasio data pelatihan dan pengujian yang bervariasi [6].

2. METODE PENELITIAN

A. Data Mining

Data mining bertujuan untuk menemukan hubungan tersembunyi, tren, atau pola yang tersembunyi dalam data yang besar dan kompleks . Data mining, seperti yang didefinisikan oleh Gartner Group, adalah proses menggunakan matematika dan statistik untuk mengungkap pola dan tren tersembunyi dalam kumpulan data yang besar [7].

B. Klasifikasi

Dalam data mining, klasifikasi berfokus pada pengelompokan titik data ke dalam kelompok yang telah ditentukan berdasarkan fitur tertentu. Ini melibatkan pembuatan model yang dapat memprediksi kategori data baru yang belum terlihat berdasarkan karakteristiknya [5]. Metode data mining klasik yang menggunakan pembelajaran mesin dikenal sebagai klasifikasi. Dengan menggunakan klasifikasi, label atau kelas dapat diberikan kepada kumpulan data yang label atau kelasnya sudah diketahui sebelumnya [8].

C. Pohon Keputusan

Pohon keputusan adalah model klasifikasi yang memprediksi kelas data menggunakan tes pada node internal dan label kelas pada node daun. Ini membandingkan nilai atribut dengan struktur pohon untuk menelusuri jalur dari akar ke daun, dan dapat dengan mudah diubah menjadi aturan klasifikasi [9].

D. Algoritma C4.5

Peneliti sering menggunakan algoritma C4.5, versi lanjutan dari ID3, yang mencakup fitur tambahan seperti penanganan nilai yang hilang, pemrosesan data kontinu, dan pemangkasan [10]. Rumus untuk menghitung nilai gain dalam algoritma klasifikasi C4.5 [11]. Untuk menyelesaikan kasus dalam algoritma C4.5, nilai dari beberapa unit perlu dicari, yaitu:

1). Entropi

Entropi sangat penting dalam ilmu data karena mengukur ketidakpastian dan keragaman informasi, yang mempengaruhi pengambilan keputusan [7]

$$\text{Entropi (S)} = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

S : Menunjukkan jumlah total sampel data

n : Mewakili jumlah partisi atau subkelompok

pi : Mewakili proporsi sampel data

2). Gain

Gain (S, A) adalah perbedaan antara entropi sebelumnya dan ketidakpastian informasi pada setiap nilai atribut. Ini mengukur seberapa baik suatu atribut mengatur dan mengklasifikasikan data. Semakin tinggi gain, semakin baik atribut tersebut memisahkan data ke dalam kelas. Algoritma C4.5 menggunakan gain untuk membangun pohon keputusan

$$Gain(S, A) = Entropy(S) - \sum_{k=0}^n * Entropy(S_i) \tag{2}$$

S : Menunjukkan jumlah total sampel data

A : atribut

n : Menunjukkan jumlah partisi

|Si| : Menjelaskan jumlah data dalam partisi ke-i

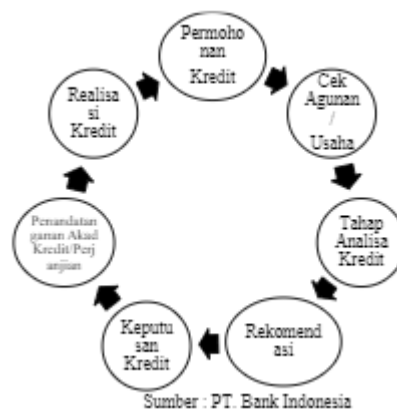
|S| : Mewakili jumlah total data dalam data S

E. Kredit Pinjaman

Kredit adalah jenis pinjaman di mana pemberi pinjaman menyediakan dana atau sumber daya kepada peminjam dengan tujuan mengembalikan jumlah yang dipinjam beserta bunga atau biaya lainnya dalam batas waktu yang ditentukan. Kredit adalah alat penting dalam kegiatan keuangan, baik untuk individu, perusahaan, maupun lembaga keuangan. Dalam Undang-Undang Perbankan No.10 Tahun 1998, perubahan atas Undang-Undang No.7 Tahun 1992, kredit adalah penyediaan uang atau tagihan yang disamakan dengan itu berdasarkan perjanjian pinjam-meminjam atau perjanjian antara bank dengan pihak lain yang mewajibkan peminjam untuk mengembalikan pinjaman [12].

F. Prosedur Pemberian Pinjaman

Pemberian pinjaman di bank memiliki tujuan yang sama yaitu menyediakan dana kepada pelanggan, tetapi penilaian kredit dan persyaratannya dapat bervariasi. Faktor-faktor seperti kebijakan bank, target pasar, dan kondisi ekonomi mempengaruhi perbedaan ini[13] . Berikut adalah alur kerja yang digunakan untuk memberikan pinjaman kredit ke bank:



Gambar 1. Aliran Kredit

Proses pemberian pinjaman bank melibatkan beberapa tahap: pelanggan mengajukan kredit, bank memeriksa jaminan atau bisnis, menganalisis faktor keuangan. Bank membuat rekomendasi persetujuan, jika disetujui, kedua belah pihak menandatangani kontrak kredit. Akhirnya, dana disalurkan kepada pelanggan sesuai dengan kesepakatan.

G. Kredit Bermasalah

Kredit macet, juga dikenal sebagai kredit bermasalah, menggambarkan situasi di mana pelanggan mengalami kesulitan atau tidak lagi mampu membayar sebagian atau seluruh kewajiban kredit mereka kepada bank [6].

H. Kolektibilitas Kredit

Peraturan Bank Indonesia No. 31/147/KEP/DIR pada 12 November 1988, untuk aset produktif bank dan tingkat kredit bermasalah ke bank dinilai berdasarkan kolektibilitas kreditnya. Seberapa baik kolektibilitas kredit seseorang menunjukkan seberapa baik mereka membayar pinjaman mereka, yang mencakup pokok, angsuran pokok, dan pembayaran bunga. Semakin baik kolektibilitas kredit seseorang, semakin besar kemungkinan bank akan mendapatkan kembali uangnya. Berdasarkan peraturan Bank Indonesia, kolektibilitas

pinjaman mencakup kredit lancar, pinjaman dalam perhatian khusus (DPK), kredit kurang lancar, kredit diragukan, dan kredit macet.

I. Penyebab Kredit Bermasalah

Saat memberikan kredit, orang menghadapi hambatan dari sumber internal dan eksternal. Hambatan internal termasuk evaluasi yang tidak akurat terhadap penggunaan kredit, tidak memahami kebutuhan keuangan pelanggan, dan kurangnya keterampilan dalam menganalisis laporan keuangan. Hambatan eksternal termasuk pemutusan hubungan kerja, penurunan ekonomi, dan kebangkrutan bisnis debitur.

J. Rapidminer

Pengembang Rapidminer di Unit Kecerdasan Buatan Universitas Dortmund adalah Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer. Ini adalah program sumber terbuka berlisensi AGPL yang digunakan untuk memproses data mining [14]. Rapidminer adalah platform sumber terbuka untuk analisis data dan penambangan pola data .

K. Python

Python adalah bahasa pemrograman populer dengan banyak keunggulan. Ini mendukung pemrograman berorientasi objek, menjadikannya ideal untuk membangun aplikasi yang kompleks dan terstruktur. Selain itu, Python dapat digunakan di berbagai platform, termasuk PC, Macintosh, dan UNIX, menjadikannya fleksibel dan dapat diakses oleh banyak pengguna [15].

L. Pembelajaran Mesin

Pembelajaran mesin adalah metode analisis data yang digunakan untuk membuat keputusan dan mengidentifikasi pola serta penyebaran dalam data sebelumnya. Ini adalah subset dari kecerdasan buatan. Sistem ini meminimalkan intervensi manusia dengan membangun model analitik otomatis 6.

M. Matriks Kebingungan

Matriks kebingungan menyediakan nilai akurasi, presisi, dan recall melalui evaluasinya, yang merupakan alat yang berguna untuk melakukan perbandingan akurasi. Akurasi klasifikasi adalah istilah untuk daftar data yang telah diklasifikasikan dengan benar setelah pengujian hasil klasifikasi.

Table 1. Confusion Matrix

	Positive	Negative
True	TP	TN
False	FP	FN

Berdasarkan nilai matriks kebingungan, akurasi, presisi, recall, dan skor F1 dapat diperoleh.

- 1) Akurasi mengukur kemampuan sistem untuk mengklasifikasikan atau memprediksi data secara tepat.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{3}$$

- 2) Presisi adalah data positif yang diklasifikasikan dengan benar dibagi dengan total.

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{4}$$

- 3) Penarikan kembali mengukur kemampuan sistem untuk mendeteksi data positif aktual.

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{5}$$

- 4) Skor F1 adalah rata-rata harmonik presisi dan ingatan.

$$F1\ Score = 2 \times \frac{presisi \times recall}{presisi+recall} \tag{6}$$

N. Teknik Resampling Teknik Oversampling Minoritas Sintetis (SMOTE)

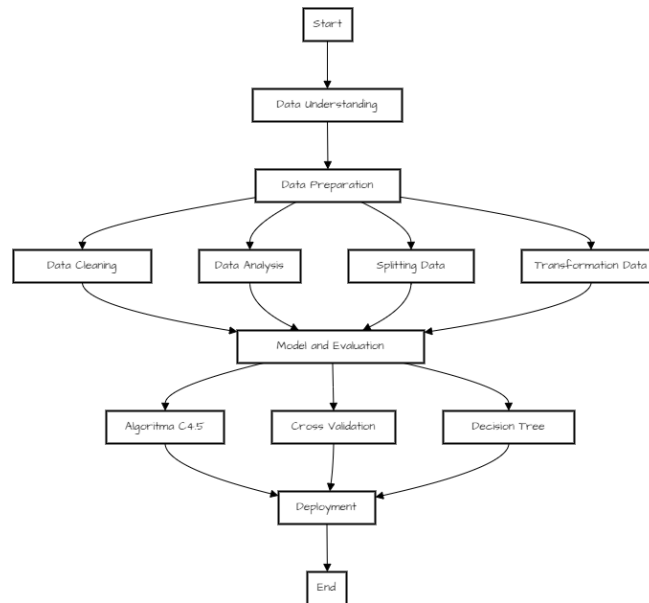
SMOTE merupakan teknik oversampling yang diusulkan oleh Chawla et al. (2002). Metode SMOTE menyeimbangkan sampel data di kelas minoritas dengan kelas mayoritas menggunakan sampel terpilih. Metode ini melibatkan perbandingan vektor fitur antar kelas [16]

O. Validasi Silang Lipatan K

K-fold cross validation merupakan salah satu model validasi silang yang digunakan untuk mengestimasi dan memprediksi akurasi model, tujuan dari k-fold cross validation adalah untuk menghilangkan informasi yang bias [17].

P. Tahapan Penelitian

Penelitian ini menggunakan metodologi standar yang disebut Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM menyederhanakan proses penambangan data untuk memecahkan masalah bisnis atau unit penelitian [18].



Gambar 2. Tahapan Penelitian

Tahap penelitian dimulai dengan pengumpulan data dari PT Menara Indonesia. Data tersebut kemudian dibersihkan, dianalisis, dan diolah. Algoritma C4.5 dibuat dan dilatih menggunakan Python dan Rapidminer. Kinerja algoritma dievaluasi dan, jika berhasil, disebar di situs web. Rincian setiap langkah adalah sebagai berikut:

1) Pemahaman data

Tahap persiapan melibatkan evaluasi kebutuhan data dan pengumpulan data yang relevan. Pemahaman mendalam tentang bisnis memandu pemilihan, sumber, dan metode pengumpulan data.

2) Persiapan data

Bagian ini menjelaskan prosedur pembersihan, analisis, pemisahan, dan transformasi data yang dilakukan untuk menyiapkan data untuk pemodelan. Pembersihan data melibatkan identifikasi dan penanganan nilai yang hilang, duplikat, dan outlier. Analisis data mencakup analisis deskriptif dan diagnostik untuk mendapatkan wawasan tentang karakteristik data. Analisis deskriptif meringkas data masa lalu menggunakan kalkulasi statistik untuk memahami perubahan dalam masalah bisnis, menjawab pertanyaan "Apa yang terjadi di masa lalu?" Di sisi lain, analisis diagnostik berusaha menemukan akar penyebab suatu masalah dengan memeriksa data atau konten. Setelah proses pengumpulan data selesai, tahap persiapan data dimulai. Pada tahap ini, data harus ditemukan, diidentifikasi, dibersihkan, dan disusun ke dalam bentuk yang diinginkan. Setelah kumpulan data dipilih, langkah berikutnya adalah memeriksa data untuk menemukan kasus yang meragukan, data yang hilang, atau data yang ambigu. Pada tahap pra-pemrosesan, data yang tidak relevan juga dibersihkan untuk mengatasi nilai yang hilang dalam data tabular. Nilai yang hilang dapat mengganggu kinerja metode yang digunakan.[19].

3) Model dan evaluasi

Pemodelan melibatkan penerapan berbagai algoritma untuk menemukan, mengidentifikasi, dan mengekspresikan pola yang tersembunyi dalam data. Setelah data siap digunakan, proses tersebut memerlukan representasi melalui model yang relevan, yang dapat memberikan pengetahuan yang berguna dan mungkin menghasilkan informasi tambahan. Model-model ini digunakan untuk menunjukkan pola dan struktur informasi, memberikan gambaran tentang tren menarik di masa mendatang. Jenis data yang digunakan menentukan model apa yang digunakan. Beberapa model umum yang digunakan meliputi estimasi, prediksi, klasifikasi, pengelompokan, dan analisis asosiasi [20]

4) Penyebaran

Pada langkah ini, peneliti menggunakan model yang dihasilkan dan menerapkannya ke sumber atau situs atau membuat laporan dan menerapkan proses penambangan data pada saat yang bersamaan. Secara ringkas, teknik CRISP-DM (Cross-Industry Standard Process for Data Mining) bertujuan untuk memberikan model bagi langkah-langkah dalam proses pencarian informasi, langkah-langkah tersebut dibagi menjadi enam yaitu: pemahaman bisnis, pemahaman data, persiapan data, pembuatan model, analisis, dan implementasi.

Q. Teknik pengumpulan data

Djaman Satori dan Aan Komariah dalam penelitiannya menjelaskan bahwa prosedur yang direncanakan secara sistematis dalam penelitian ilmiah untuk pengumpulan data [21]. Dalam penelitian yang dilakukan, data yang dikumpulkan adalah data primer yang diperoleh langsung dari perusahaan pada saat magang.

R. Metode Penelitian

Metode penelitian yang digunakan peneliti adalah kuantitatif dengan menggunakan algoritma C4.5 menggunakan alat bantu seperti Python dan Rapidminer. Pembuatan algoritma C4.5 juga melakukan beberapa percobaan seperti menggunakan oversampling dan cross-validation. Kemudian, peneliti akan mengevaluasi hasil keseluruhan dan membandingkannya dengan pembuatan algoritma C4.5 tanpa menggunakan oversampling atau cross-validation.

3. HASIL DAN ANALISIS

Proses klasifikasi pada data dengan algoritma C4.5 pada dataset yang telah disiapkan untuk memprediksi kelayakan kredit nasabah dengan bantuan Rapidminer dan Python.

A. Pemahaman Data

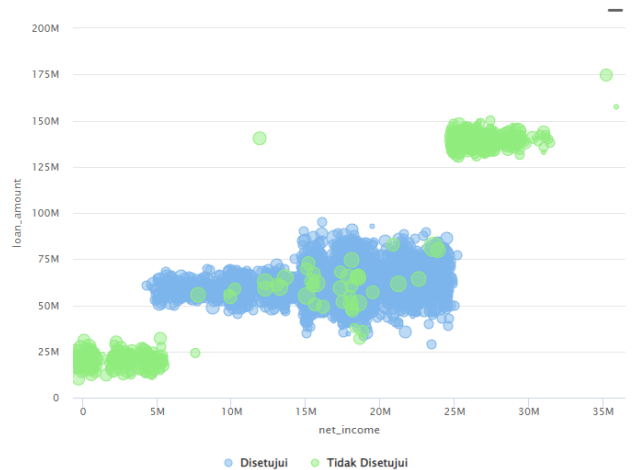
Dataset yang digunakan dalam penelitian ini adalah `net_income`, `loan_amount`, `tenor`, `arrears_amounts`, `assets`, `credit score`, `ageing`, dan `credit_risk`. Berikut ini adalah rinciannya.

Variabel	Tipe Data	Keterangan
pendapatan bersih	Bilangan bulat	Pendapatan bersih pelanggan.
jumlah_pinjaman	Bilangan bulat	Pengajuan tagihan debitur atau pelanggan.
penyanyi tenor	Bilangan bulat	Waktu pembayaran oleh debitur atau nasabah.
jumlah_tunggakan	Bilangan bulat	Tunggakan cicilan pelanggan.
aktiva	Bilangan bulat	Jumlah tabungan yang tersisa yang dimiliki nasabah.
skor_kredit	Bilangan bulat	Skor kredit bank nasabah.
penuaan	Kategori	Status penuaan pelanggan
risiko kredit	Kategori	Status persetujuan pelanggan.

Ini adalah kumpulan data yang telah disiapkan yang terdiri dari 5000 baris dan 8 kolom. Ada juga definisi tambahan tentang penuaan. Penuaan adalah durasi tunggakan pelanggan dalam industri pembiayaan modal..

B. Persiapan Data

Setelah pengumpulan data, proses persiapan data yang komprehensif dilakukan untuk memastikan kualitas dan kesesuaian data untuk pemodelan. Proses ini melibatkan pembersihan data, analisis, pemisahan, dan transformasi. Pembersihan data memerlukan pemeriksaan menyeluruh terhadap kumpulan data untuk mengidentifikasi dan menghilangkan nilai yang hilang, duplikat, atau ketidakkonsistenan. Hal ini menghasilkan kumpulan data yang bersih tanpa masalah kualitas data. Selanjutnya, para peneliti melakukan analisis data yang komprehensif, menggunakan teknik deskriptif dan diagnostik. Berikut adalah hasil analisisnya.



Gambar 3. Analisis Multivariat

Peneliti menganalisis laba bersih, jumlah pinjaman, risiko kredit, dan jumlah tunggakan. Hasilnya menunjukkan bahwa nasabah dengan pendapatan ekstrem sering kali tidak disetujui oleh pemberi pinjaman. Selain itu, data laba bersih rata-rata ditolak karena jumlah tunggakan yang sedikit lebih tinggi. Temuan utama dari analisis tersebut mengungkapkan adanya outlier, variasi skala yang signifikan di seluruh variabel, dan hubungan timbal balik yang kompleks di antara variabel. Selain itu, kumpulan data tersebut menunjukkan ketidakseimbangan kelas, dengan 4.372 dan 628 contoh di kelas yang disetujui dan tidak disetujui. Untuk mengatasi ketidakseimbangan ini, Teknik Synthetic Minority Oversampling (SMOTE) digunakan. Akhirnya, data dibagi menjadi set pelatihan dan pengujian dan ditransformasikan untuk pelatihan dan evaluasi model.

C. Model dan Evaluasi

Setelah data disiapkan, berbagai model dikembangkan dan dievaluasi menggunakan algoritma C4.5 di lingkungan Python dan RapidMiner. Dalam Python, kinerja model dengan dan tanpa oversampling dibandingkan. Peneliti membuat model atau algoritma C4.5 atau yang biasa disebut Decision Tree dan dilatih dengan data pelatihan yang telah dibuat sebelumnya sebagai berikut.

```
tree1 = DecisionTreeClassifier(max_depth=6, random_state=42)
tree1.fit(X_train, y_train)
✓ 0.0s
DecisionTreeClassifier
DecisionTreeClassifier(max_depth=6, random_state=42)
```

Gambar 4. Membangun Model dengan Python

Setelah model dibuat, para peneliti mengevaluasinya menggunakan akurasi, presisi, recall, dan f1-score. Berikut hasilnya:

Tabel 2. Perbandingan Matrik Evaluasi

Metrik Evaluasi	Tanpa SMOTE(%)	SEMUT (%)	Meningkatkan (%)
Ketepatan	99,40%	99,70%	0,30%
Presisi	99,24%	99,25%	0,01%
Mengingat	96,30%	98,52%	2,22%
F1-Skor	97,74%	98,88%	1,14%

Hasilnya menunjukkan peningkatan signifikan dalam performa model saat oversampling digunakan. Selanjutnya, perbandingan performa model antara implementasi Python dan Rapidminer. Berikut ini adalah hasil implementasi di Rapidminer.

● Table View ○ Plot View

accuracy: 99.70%

	true Disetujui	true Tidak Disetujui	class precision
pred. Disetujui	873	2	99.77%
pred. Tidak Disetujui	1	124	99.20%
class recall	99.89%	98.41%	

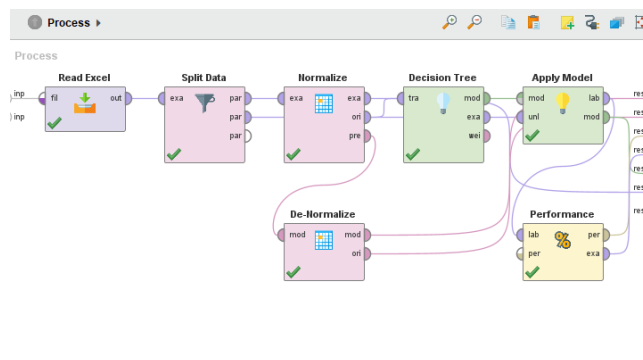
Gambar 5. Membangun Model dengan Rapidminer

Perbandingan performa model antara implementasi Python dan Rapidminer menunjukkan performa yang lebih unggul untuk algoritma C4.5 berbasis Python. Namun, perbedaan performa antara kedua pendekatan tersebut tidak terlalu signifikan. Untuk mengevaluasi performa model lebih lanjut, dilakukan validasi silang k-fold, yang menghasilkan hasil yang memuaskan. Berikut hasilnya:

Tabel 3. Validasi Silang

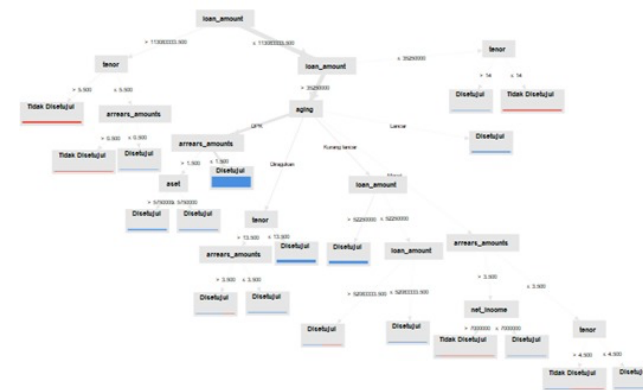
<i>Lipatan</i>	<i>Ketepatan</i>
1	98,97%
2	99,09%
3	99,66%
4	98,97%
5	99,54%
6	99,43%
7	99,20%
8	99,54%
9	99,66%
10	99,31%

Setiap lipatan tidak menunjukkan perbedaan yang signifikan sehingga akurasi rata-ratanya adalah 99,34%. Setelah kemampuan model untuk memprediksi data secara akurat dikonfirmasi, pohon keputusan dibuat menggunakan Rapidminer untuk membuat struktur pohon berdasarkan fitur yang ada.



Gambar 6. Pohon Keputusan di Rapidminer

Hasil pohon keputusan yang telah dibuat dengan menggunakan desain tersebut adalah sebagai berikut.

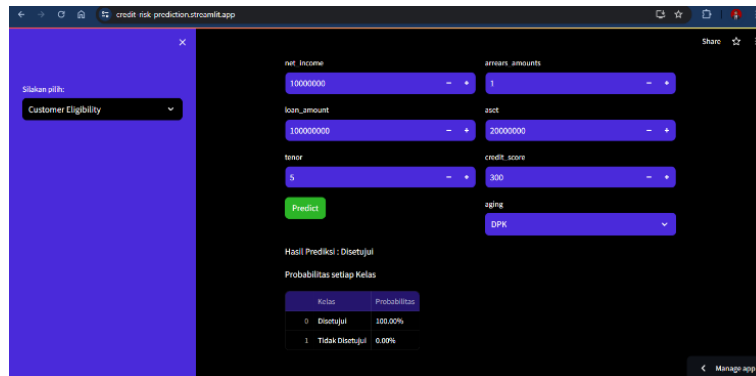


Gambar 7. Pohon Keputusan

Pohon keputusan yang menghasilkan 19 aturan dengan "loan_amount" sebagai simpul akar. Artinya, loan_amount mempunyai perolehan informasi tertinggi dan dianggap sebagai fitur paling penting.

D. Penyebaran

Pada tahap ini merupakan tahap implementasi hasil akhir dari proyek penelitian. Pada tahap ini, peneliti mencoba menguji model tersebut sebagai model prediktif yang telah dikembangkan. Tahap ini menggunakan website dashboard Streamlit sebagai interface sederhana untuk mengimplementasikan sistem prediksi kelayakan nasabah. Website ini dapat diakses di <https://credit-risk-prediction.streamlit.app/>. Proses ini dapat dilihat pada gambar di bawah ini.



Gambar 8. Penerapan pada Streamlit

Hasil prediksi menunjukkan disetujui dengan probabilitas atau kemungkinan disetujui sebesar 100%.

4. KESIMPULAN.

Penelitian ini telah menghasilkan sejumlah temuan penting yang dapat digabungkan menjadi suatu simpulan yang komprehensif. Metode algoritma C4.5 dalam memprediksi kelayakan nasabah memberikan prediksi yang akurat dan dapat membantu pihak pemberi pinjaman dalam menentukan kelayakan nasabah secara lebih cepat dan efisien dibandingkan dengan menggunakan metode tradisional, sehingga terciptanya algoritma ini sangat berpengaruh terhadap prediksi kelayakan nasabah secara cepat, tepat, dan efisien. Membangun model klasifikasi dengan algoritma C4.5 untuk memprediksi kelayakan kredit nasabah di PT Menara Indonesia dapat dicapai dengan berbagai cara, salah satunya menggunakan bahasa pemrograman Python dan perangkat lunak penambangan data RapidMiner. Integrasi algoritma C4.5 membuktikan keberhasilannya dalam memprediksi kelayakan kredit nasabah yang dibuktikan dengan hasil akurasi yang tinggi dan kinerja model yang baik di atas 99%. Implementasi atau deployment yang dilakukan memudahkan peneliti dalam menginterpretasikan hasil prediksi yang kemungkinan hasil model tersebut juga dapat memudahkan pihak pemberi pinjaman dalam memprediksi kelayakan nasabah.

REFERENSI

- [1] K. Auliasari, M. Kertaningtyas, and D. Wilis Lestaring Basuki, "Analisis Penentuan Resiko Kredit Menggunakan Algoritma C.5.0," *J-TIT*, vol. 8, no. 1, Jun. 2021, doi: 10.25047/jtit.v8i1.218.
- [2] D. Marlina and M. Bakri, "Penerapan Data Mining Untuk Memprediksi Transaksi Nasabah Dengan Algoritma C4.5," *Jurnal Teknologi Dan Sistem Informasi*, vol. 2, no. 1, pp. 23–28, 2021.
- [3] E. Edward, "Identifikasi Berita HOAX Berbasis Web Menggunakan Algoritma C4.5," *jiksi*, vol. 9, no. 1, p. 53, Jan. 2021, doi: 10.24912/jiksi.v9i1.11558.
- [4] N. Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *Jurnal Teknik Komputer AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022.
- [5] T. Novianti, S. A. Mandati, and E. K. Andana, "Peningkatan Evaluasi Risiko Kredit Menggunakan Decision Tree C 4.5," *Mine-tech*, vol. 2, no. 2, pp. 1–9, Dec. 2023, doi: 10.30651/mine-tech.v2i2.21749.
- [6] T. Hidayatulloh, A. Fajria, R. N. Lestari, and N. S. Z. Nufus, "Algoritma C4.5 Untuk Menentukan Kelayakan Pemberian Kredit (Studi kasus: Bank Mandiri Taspen Kantor Kas Sukabumi)," *LTK*, vol. 2, no. 2, pp. 66–74, Dec. 2022, doi: 10.31294/larik.v2i2.1836.
- [7] D. Puspita, S. Aminah, and A. Arif, "Prediction System for Credit Eligibility Using C4.5 Algorithm," *JITE*, vol. 6, no. 1, pp. 148–156, Jul. 2022, doi: 10.31289/jite.v6i1.7311.
- [8] D. Sartika and I. Saluza, "Penerapan Metode Principal Component Analysis (PCA) Pada Klasifikasi Status Kredit Nasabah Bank Sumsel Babel Cabang KM 12 Palembang Menggunakan Metode Decision Tree," *generic*, vol. 14, no. 2, pp. 45–49, Jul. 2022, doi: 10.18495/generic.v14i2.130.
- [9] F. F. Kusuma, "Penerapan Data Mining Untuk Akurasi Analisis Cuaca di Australia Menggunakan Algoritma J48 Decision Tree," *j.inf.syst.int.*, vol. 3, no. 2, pp. 65–68, Jul. 2023, doi: 10.53514/jco.v3i2.396.
- [10] N. Handayani, H. Wahyono, J. Trianto, and D. S. Permana, "Prediksi Tingkat Risiko Kredit dengan Data Mining Menggunakan Algoritma Decision Tree C.45," *Jur. Ris. Kom.*, vol. 8, no. 6, p. 198, Dec. 2021, doi: 10.30865/jurikom.v8i6.3643.
- [11] R. Rofiani, L. Oktaviani, D. Vernanda, and T. Hendriawan, "Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5," *Jurnal Tekno Kompak*, vol. 18, no. 1, pp. 126–139, 2024.
- [12] R. Oktapiani, D. Prayudi, A. Fajria, N. S. Z. Nufus, and R. N. Lestari, "Sistem Pendukung Keputusan Untuk Menentukan Manajemen Kelayakan Pemberian Kredit Di Bank Mandiri Taspen Sukabumi Menggunakan Metode Analytic Hierarchy Process," *IJSE*, vol. 8, no. 1, pp. 36–45, Jun. 2022, doi: 10.31294/ijse.v8i1.12054.
- [13] R. Widayati and U. Herman, "PENYELESAIAN KREDIT BERMASALAH PADA PT. BANK PERKREDITAN RAKYAT (BPR) NAGARI KASANG," Feb. 28, 2019, *Open Science Framework*. doi: 10.31219/osf.io/d738z.
- [14] L. Fimawahib and E. Rouza, "Penerapan K-Means Clustering pada Penentuan Jenis Pembelajaran di Universitas Pasir Pengaraian," *ISI*, vol. 6, no. 2, p. 234, Nov. 2021, doi: 10.35314/isi.v6i2.2096.
- [15] T. Novianti, S. A. Mandati, and E. K. Andana, "Peningkatan Evaluasi Risiko Kredit Menggunakan Decision Tree C 4.5," *MINE-TECH: Journal of Manufacturing in Industrial Engineering & Technology*, vol. 2, no. 2, pp. 1–9, 2023.
- [16] C. Mulia and A. Kurniasih, "Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Bank Customer Churn Menggunakan Algoritma Naïve bayes dan Logistic Regression," in *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, 2023, pp. 552–559.
- [17] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 6, no. 9, pp. 4305–4313, 2022.
- [18] D. Wibisono, "Pengukuran Kualitas Sistem Informasi Kasir Terhadap Layanan Konsumen Menggunakan Metode Crisp-Dm Variabel Servqual (Studi Kasus: Studio Foto Archa)," *Jurnal Teknologi dan Sistem Informasi*, vol. 3, no. 4, pp. 112–119, 2022.
- [19] A. N. Puteri, A. Arizal, and A. D. Achmad, "Feature Selection Correlation-Based pada Prediksi Nasabah Bank Telemarketing untuk Deposito," *matrik*, vol. 20, no. 2, pp. 335–342, May 2021, doi: 10.30812/matrik.v20i2.1183.
- [20] A. Prasetyo, R. Sastra, and N. Musyaffa, "IMPLEMENTASI DATA MINING UNTUK ANALISIS DATA PENJUALAN DENGAN MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS DAPOERIN'S)," *JKI*, vol. 8, no. 2, Dec. 2020, doi: 10.31294/jki.v8i2.8994.
- [21] A. Febriyani, G. K. Prayoga, and O. Nurdiawan, "Index Kepuasan Pelanggan Informa dengan Menggunakan Algoritma C.45," *Jur. Ris. Kom.*, vol. 8, no. 6, p. 330, Dec. 2021, doi: 10.30865/jurikom.v8i6.3686.