

# Implementasi Algoritma C4.5 Melalui Pohon Keputusan (Decision Tree) berbasis Metode Forward Selection Untuk Memprediksi Risiko Kredit Macet

<sup>1</sup>Ahmad Yani, <sup>2</sup>Febriyansyah Ramadhan, <sup>3</sup>Debi Irawan, <sup>4</sup>Asep Wasid  
<sup>1,2,3,4</sup>Universitas Indonesia Membangun  
Jakarta, Indonesia

<sup>1</sup>eahmadyani.tfi@student.inaba.ac.id, <sup>2</sup>febriansyah.ramadhan@inaba.ac.id,  
<sup>3</sup>debi.irawan@inaba.ac.id, <sup>4</sup>asep.wasid@inaba.ac.id

\*Penulis Korespondensi

Diajukan : 27/10/2025  
Diterima : 31/10/2025  
Dipublikasi : 31/10/2025

## ABSTRAK

Perusahaan pembiayaan di era digital yang dengan cepat mengadopsi teknologi informasi, telah mengubah perilaku konsumen untuk lebih mengutamakan pengalaman seamless dan transaksi mandiri melalui aplikasi mobile. Seringkali tanpa penilaian risiko yang cukup, serta permintaan akan layanan real-time, yang secara keseluruhan meningkatkan potensi terjadinya risiko kredit macet. Untuk mengatasi isu yang ada, peneliti melaksanakan sebuah studi di PT Trust Finance Indonesia guna memproyeksikan risiko gagal bayar. Riset ini memanfaatkan algoritma C4.5 yang dikombinasikan dengan metode *forward selection*. Adapun fokus penelitian ini adalah untuk menganalisis kapabilitas metode yang diusulkan sebagai alat estimasi risiko kredit macet. Sebanyak 300 set data diambil dari PT Trust Finance Indonesia sebagai sampel penelitian. Proses selanjutnya adalah pengolahan data menggunakan aplikasi *RapidMiner*. Pemrosesan ini menguji dua skenario berbeda, yaitu algoritma C4.5 dasar dan algoritma C4.5 yang telah disempurnakan dengan metode *forward selection*. Hasil pengujian C4.5 tanpa peningkatan memperoleh rata-rata keakuratan sebesar 83,33%, class recall sebesar 88,24% true lancar, 76,02% true berisiko macet. Hasil pengujian C4.5 dengan peningkatan memperoleh rata-rata keakuratan 87,59%, tergolong dalam kategori klasifikasi yang baik

**Kata Kunci:** Algoritma C4.5, Data Mining, Klasifikasi, Kredit

## I. PENDAHULUAN

Kemajuan pesat di bidang teknologi informasi telah memicu perubahan signifikan dalam beragam aspek kehidupan, yang juga mencakup sektor bisnis keuangan. PT Trust Finance Indonesia merupakan salah satu perusahaan di bidang keuangan yang berfokus pada penyaluran dana kepada nasabah untuk keperluan pengadaan barang atau jasa. Berbeda dengan bank, perusahaan ini tidak menghimpun dana publik dan operasionalnya diawasi oleh Otoritas Jasa Keuangan (OJK). Kegiatan bisnisnya terdiri dari pembiayaan konsumen seperti kredit kendaraan, sewa guna usaha (leasing), serta anjak piutang (factoring). Di tengah persaingan yang semakin sengit, perusahaan ini berkomitmen untuk meningkatkan layanan dan efisiensi dalam pengambilan keputusan, khususnya terkait pemberian kredit (Ita Kumaratih, 2017). Mengingat banyaknya nasabah yang memerlukan akses dana untuk menunjang pertumbuhan usaha mereka, maka mengadopsi teknologi melalui aplikasi mobile. Secara tidak langsung mengubah perilaku konsumen untuk lebih mengutamakan pengalaman seamless dan transaksi mandiri secara real-time. Seringkali tanpa analisa risiko yang memadai secara keseluruhan meningkatkan risiko kredit macet (Arie Yandi Saputra, 2021).

Risiko Kredit didefinisikan sebagai potensi kerugian yang dialami suatu entitas karena



mitra pengimbang (pihak lain) gagal memenuhi kewajibannya. Risiko ini dapat diklasifikasikan lebih lanjut menjadi beberapa jenis. Di antaranya adalah risiko yang murni disebabkan oleh ketidakmampuan debitur, risiko akibat pemusatan dalam alokasi dana (Risiko Konsentrasi Kredit), dan risiko yang timbul dari kegagalan mitra transaksi (risiko kredit pihak lawan). Kategori lainnya mencakup risiko terkait kegagalan proses penyelesaian transaksi (risiko penyelesaian) serta risiko yang dipicu oleh faktor kondisi suatu negara (risiko negara). Sebuah pinjaman dianggap bermasalah saat peminjam gagal mengembalikan pinjaman pokok beserta bunganya, atau salah satunya, sesuai dengan jadwal yang telah ditentukan. Kegagalan ini bisa berupa keterlambatan pembayaran ataupun ketidakmampuan untuk membayar. Di sektor perbankan, kredit yang bermasalah sering diistilahkan sebagai Non-Performing Loan (NPL, yakni pinjaman yang tidak lagi memberikan keuntungan bagi pemberi pinjaman. Dalam menangani isu kredit bermasalah, penggunaan sistem penilaian yang objektif dan berbasis data sangatlah penting) (Riyandi Nur Sumawidjaya, 2015).

Metode penilaian algoritma C4.5 yang didasarkan pada Forward Selection untuk mengklasifikasikan nasabah dapat memberikan dukungan bagi perusahaan pembiayaan dalam meningkatkan pengelolaan risiko kredit dengan lebih efektif. Dengan menggunakan fitur-fitur masukan seperti jumlah pinjaman, suku bunga, dan jangka waktu pinjaman, algoritma ini mampu memberikan ramalan yang lebih tepat mengenai kemungkinan seorang nasabah membayar pinjamannya tepat waktu. Ini akan memudahkan perusahaan pembiayaan dalam membuat keputusan yang lebih baik terkait pemberian pinjaman dan pengelolaan risiko kredit secara keseluruhan (Bentar Candra P, 2024).

Sebuah penelitian yang dilakukan oleh (Adi, 2015) berfokus pada prediksi pinjaman bermasalah di koperasi simpan pinjam dengan memanfaatkan perilaku nasabah. Dalam studinya, Sucipto menerapkan algoritma klasifikasi C4.5 dan menguji model tersebut menggunakan metode ROC/AUC serta T-Test. Hasil penelitian ini memperlihatkan performa yang sangat baik, dengan perolehan akurasi 91,06%, presisi 100%, dan recall 78%. Temuan ini mengindikasikan bahwa algoritma C4.5 secara umum efektif untuk dijadikan referensi dalam mendeteksi risiko gagal bayar sebelum pinjaman disetujui.

PT Trust Finance Indonesia memiliki tanggung jawab dalam mengelola risiko, keputusan untuk memberikan pinjaman bukanlah sesuatu yang dapat dianggap sepele. Keputusan mengenai kelayakan kredit nasabah masih diambil oleh pimpinan berdasarkan naluri dan pengalaman mereka. Hasil survei kepada calon nasabah tidak dapat dijadikan patokan utama karena kenyataannya di lapangan menunjukkan bahwa hampir semua nasabah sudah mengikuti proses survei dan dinyatakan layak mendapatkan pinjaman, tetapi masih ada beberapa nasabah yang mengalami masalah kredit. Sehingga menimbulkan masalah pada analisa kredit yang tidak tepat, yang dapat berakibat pada berbagai risiko, seperti kehilangan nasabah, ketidakpastian dalam pembayaran pinjaman, bahkan nasabah yang tidak dapat melunasi pinjamannya (Menarianti, 2015). Penelitian ini bertujuan untuk menelaah efektivitas penerapan algoritma C4.5 sebagai metode untuk memprediksi risiko kredit bermasalah nasabah di PT Trust Finance Indonesia. Oleh karena itu, studi ini mengimplementasikan algoritma Klasifikasi C4.5 yang dikombinasikan dengan decision tree (pohon keputusan) berbasis forward selection. Metode gabungan ini digunakan untuk prediksi risiko gagal bayar di PT Trust Finance Indonesia. Alasan pemilihan metode tersebut adalah ketersediaan tingkat akurasi yang terbukti memadai untuk kebutuhan prediksi dan klasifikasi data.

## II. STUDI LITERATUR

### Penelitian Terdahulu

Sebagai landasan bagi penelitian yang sedang dilaksanakan, seorang peneliti diwajibkan untuk mengkaji, menganalisis, serta menyajikan beberapa studi relevan sebelumnya. Dalam konteks ini, disyaratkan penggunaan paling sedikit tiga penelitian terdahulu yang dipublikasikan dalam kurun waktu lima tahun terakhir.

#### 1. Algoritma C4.5

Algoritma C4.5 adalah salah satu metode yang diklasifikasikan ke dalam keluarga algoritma pohon keputusan. Dalam implementasinya, C4.5 berfungsi untuk mengkonstruksi sebuah model sistem yang hasil akhirnya adalah sebuah pohon keputusan. Pohon ini selanjutnya mampu

menghasilkan aturan-aturan yang digunakan untuk memprediksi risiko kredit. Efektivitas C4.5 dalam klasifikasi data dianggap sangat tinggi. Tingginya efektivitas ini disebabkan oleh akurasi yang dapat dicapai saat memperoleh informasi dari data yang terklasifikasi, sehingga memberikan kemudahan bagi peneliti untuk mendapatkan informasi dari data tersebut. Kemampuan klasifikasi C4.5 yang tinggi juga didukung oleh fitur data yang teridentifikasi dengan jelas. Fitur-fitur dari data yang sudah diidentifikasi secara gamblang dapat ditampilkan menggunakan format pohon keputusan (decision tree) ataupun kaidah if-then. Metode penyajian ini bertujuan agar pemakai lebih mudah dalam memahami dan melihat informasi yang terkait dalam penelitian ini sangat bergantung pada pemilihan atribut. Oleh karena itu, proses seleksi atribut memiliki peranan yang sangat vital, sebab hasil akhir yang didapat sangat dipengaruhi oleh atribut yang dipilih (Eka Praja Wiyata Mandala, 2018).

Sebagaimana telah diuraikan, algoritma C4.5 dimanfaatkan dalam penelitian ini untuk memfasilitasi peramalan nilai informasi. Algoritma ini dipandang sebagai penyempurnaan dari metode ID3. Keduanya digunakan untuk menghasilkan pohon keputusan yang dapat menilai atau meramalkan sebuah kejadian. Tujuan akhirnya adalah untuk mendapatkan informasi yang esensial. Adapun langkah-langkah operasional dalam Algoritma C4.5 adalah sebagai berikut (Ardiyansyah, 2023):

- a. Penentuan nilai gain. Proses ini mengharuskan pemilihan atribut pada masing-masing node dalam pohon keputusan. Aktivitas ini krusial untuk mengidentifikasi variabel yang akan dijadikan node pada pohon keputusan itu. Formula berikut digunakan untuk mengkalkulasi nilai gain::

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S) \quad (1)$$

Dalam konteks ini, S adalah himpunan yang menampung seluruh kasus, dan A mewakili atribut yang sedang dianalisis. Atribut A ini dibagi menjadi n partisi (pembagian). Simbol  $|S_i|$  merujuk pada jumlah kasus yang termasuk dalam partisi ke-i, sementara  $|S|$  adalah notasi untuk jumlah total dari semua kasus dalam himpunan S.

- b. Selanjutnya, dilakukan penentuan nilai entropi. Fungsi dari nilai ini adalah untuk mengevaluasi tingkat variasi pilihan yang ada pada sebuah nilai tertentu. Perhitungan tersebut dilakukan dengan menggunakan formula di bawah ini:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (2)$$

Dalam formula ini, S didefinisikan sebagai keseluruhan himpunan data (kasus), sedangkan A adalah atribut yang digunakan. Variabel n melambangkan jumlah total bagian dari atribut A. Sementara itu,  $P_i$  merepresentasikan rasio atau proporsi dari subset  $S_i$  terhadap keseluruhan himpunan S.

## 2. Metode Forward Selection

Sebuah dataset acapkali memiliki jumlah fitur yang sangat banyak, bahkan bisa mencapai ratusan. Permasalahannya, tidak semua fitur tersebut memiliki kaitan atau relevansi dengan pola yang sedang diteliti. Sebagai contoh, dalam kasus klasifikasi untuk menentukan status kredit seorang nasabah (apakah baik atau bermasalah), variabel seperti nomor telepon tentu memiliki tingkat kepentingan yang jauh lebih rendah dibandingkan variabel usia atau riwayat kreditnya. Apabila atribut-atribut yang tidak relevan ini tetap dipertahankan, proses penemuan pola dapat menjadi tidak efisien. Hal ini juga berpotensi mengacaukan kinerja algoritma data mining, sehingga pola yang ditemukan pada akhirnya memiliki kualitas yang buruk. Salah satu metode untuk menangani permasalahan ini adalah dengan melakukan eliminasi atau pembuangan terhadap atribut-atribut yang dianggap tidak penting tersebut.

Forward Selection didefinisikan sebagai sebuah metode dalam Pemilihan Fitur (Feature Selection). Tujuan utamanya adalah untuk mengidentifikasi himpunan fitur (subset) yang paling ringkas namun tetap optimal dari sekumpulan besar atribut. Penggunaan teknik Pemilihan Fitur ini bermanfaat untuk menghasilkan model atau pola yang lebih sederhana dan mudah diinterpretasi. Alur kerja metode ini berawal dari ketiadaan fitur (himpunan kosong). Selanjutnya,

dilakukan evaluasi terhadap setiap atribut secara satu per satu. Atribut yang terbukti memberikan kontribusi akurasi tertinggi akan ditambahkan ke dalam himpunan fitur terpilih. Pengujian ini bersifat iteratif dan akan dihentikan ketika penambahan atribut manapun tidak lagi memberikan perbaikan akurasi yang substansial. (Lusi Asameta, 2024):

- a. {A1}
- b. {A1, A4}
- c. Hasil pengurangan atribut: {A1, A4, A6} Confusion Matrix

### 3. Matriks Confusion

*Matriks confusion* adalah alat evaluasi kerja yang penting dalam klasifikasi *data mining*, memberikan gambaran menyeluruh tentang hasil prediksi model. Dalam Mengklasifikasikan biner dibagi menjadi dua kategori, misalnya kategori 0 dan 1. Dalam konteks evaluasi model, *Confusion Matrix* membagi hasil prediksi menjadi empat kategori. Kategori pertama adalah *True Positive* (TP), yang merujuk pada data yang kondisi aktualnya positif dan berhasil diprediksi dengan benar sebagai positif. Kategori kedua adalah *True Negative* (TN), yaitu data yang kondisi aktualnya negatif dan juga berhasil diprediksi dengan benar sebagai negatif. Terdapat pula dua skenario kesalahan: *False Positive* (FP) terjadi ketika data yang aktualnya negatif namun keliru diprediksi sebagai positif, serta *False Negative* (FN) yang terjadi saat data yang aktualnya positif justru keliru diprediksi sebagai negative (Anton Saputra, 2022):

#### a. Akurasi

Ini adalah persentase dari data pengujian yang dengan tepat dikategorikan oleh sistem klasifikasi. Merupakan proporsi dari himpunan tupel pengujian yang berhasil dikategorikan dengan tepat oleh pengklasifikasi.

$$\text{Akurasi} = \frac{\text{TP} - \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

#### b. Sensitivitas

Ini adalah persentase dari data pengujian yang dengan tepat dikategorikan oleh sistem klasifikasi. Merupakan proporsi dari himpunan tupel pengujian yang berhasil dikategorikan dengan tepat oleh pengklasifikasi.

$$\text{Sensitivitas} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

#### c. Spesifisitas

Spesifisitas, yang juga dikenal dengan istilah *True Negative Rate* (Tingkat Negatif Sebenarnya), mengacu pada proporsi dari total instans negatif yang prediksinya benar (diklasifikasikan sebagai negatif).

$$\text{Spesifisitas} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

#### d. PPV

Adalah persentase kasus yang memiliki hasil tes positif dan didiagnosis secara akurat negatif).

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

#### e. NPV

Adalah persentase kasus yang memiliki hasil tes positif dan didiagnosis secara akurat.

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (7)$$

### 4. Area Under ROC Curve

Sebuah grafik yang dikenal sebagai Kurva ROC (*Receiver Operating Characteristic*) berfungsi untuk memvisualisasikan relasi antara dua metrik. Metrik tersebut adalah sensitivitas, yang dipetakan pada sumbu Y, dan nilai '1 - spesifisitas' (atau *False Positive Rate*), yang dipetakan

pada sumbu X.

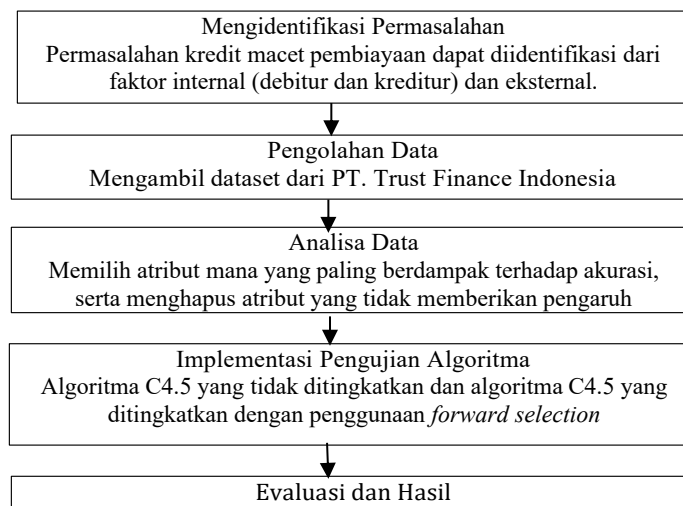
Nilai yang didapat dari Kurva ROC (sering disebut AUC atau *Area Under Curve*) dapat dimanfaatkan sebagai sebuah tolok ukur evaluasi. Tolok ukur ini berguna untuk melakukan komparasi performa antara beberapa algoritma yang berbeda. Melalui kurva ROC, proses klasifikasi dapat divisualisasikan, diatur, serta ditentukan berdasarkan capaian kinerja dari algoritma tersebut.

Dalam data mining, nilai AUC (*Area Under ROC Curve*) dimanfaatkan sebagai metrik untuk mengevaluasi keakuratan sebuah model klasifikasi. Kinerja model tersebut dapat diinterpretasikan ke dalam lima tingkatan berbeda:

- 0.90 - 1.00 = Menunjukkan performa klasifikasi yang sangat baik (istimewa)
- 0.80 - 0.90 = Menunjukkan performa klasifikasi yang baik.
- 0.70 - 0.80 = Menunjukkan performa klasifikasi yang cukup baik (sedang).
- 0.60 - 0.70 = Menunjukkan performa klasifikasi yang kurang baik (lemah)
- 0.50 - 0.60 = Menunjukkan performa klasifikasi yang tidak baik (buruk).

### III. METODE

Dalam penelitian ini diterapkan metode klasifikasi C4.5, di mana klasifikasi merupakan langkah untuk mengidentifikasi sekumpulan pola atau fungsi yang menggambarkan serta memisahkan satu kelas data dari kelas data lainnya. Berikut adalah langkah-langkah dalam proses penelitian mengenai perkiraan risiko kredit di PT. Trust Finance Indonesia yang dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

Mengacu pada Gambar 1, studi ini mencakup enam tahapan yang sistematis. Tahapan-tahapan tersebut meliputi Identifikasi Masalah, Akuisisi Data, Analisis Data, Implementasi Algoritma, Pengujian, serta Evaluasi dan Hasil. Penjelasan lebih rinci untuk setiap tahapan akan dipaparkan pada bagian berikut:

- Mengidentifikasi Permasalahan**  
Sebelum memulai sebuah penelitian, tahap pertama yang krusial adalah mengidentifikasi masalah. Penelitian ini akan fokus pada cara untuk memperkirakan kemungkinan terjadinya kredit macet di PT Trust Finance Indonesia.
- Pengolahan Data**  
Mengumpulkan dataset nasabah PT Trust Finance Indonesia yang akan dimanfaatkan untuk keperluan penelitian.
- Analisa Data**  
Mengkaji informasi data yang telah dikumpulkan dan kemudian disaring dengan metode *forward selection*, dengan cara memilih fitur yang memiliki dampak terbesar dan menyingkirkan fitur yang tidak berpengaruh, kemudian dibuatlah pohon keputusan.

## d. Implementasi Pengujian algoritma

Pada fase ini, pengkodean sistem dilakukan sesuai dengan rencana yang telah disusun. algoritma C4.5 diterapkan sebagai komponen dari modul prediksi dalam sistem ini. Model dari algoritma C4.5 akan memanfaatkan dataset untuk memperkirakan risiko kredit yang mungkin tidak terbayar. Seluruh elemen sistem, termasuk penggabungan dengan algoritma C4.5, dikembangkan dengan cara yang terstruktur. Hasil dari analisa pengujian algoritma dilakukan evaluasi dengan penerapan algoritma C4.5 melalui Metode Forward Selection.

## e. Evaluasi dan Hasil

Untuk mengevaluasi kinerja dan memverifikasi fungsionalitas model yang telah dikembangkan, akan dilaksanakan analisis terhadap data hasil pengujian. Penelitian ini menggunakan confusion matrix sebagai tolok ukur utama untuk melakukan evaluasi. Dari matriks tersebut, diperoleh berbagai metrik perhitungan seperti presisi, recall, dan akurasi. Nilai-nilai ini didapatkan berdasarkan komponen-komponen penyusun confusion matrix (yaitu true positive, false positive, true negative, dan false negative). Perhitungan ini bertujuan membandingkan performa antara dua varian algoritma C4.5: versi yang dioptimalkan dan versi yang tidak dioptimalkan. Secara definitif, presisi mengukur kesesuaian antara informasi yang diberikan sistem dengan kueri pengguna. Berbeda dengan itu, recall menyoroti kapasitas sistem untuk menemukan kembali seluruh informasi yang relevan. Adapun akurasi merepresentasikan tingkat kesamaan antara hasil prediksi dan nilai faktualnya.

#### IV. HASIL DAN PEMBAHASAN

Dari penelitian dan identifikasi permasalahan yang sudah dilakukan, terdapat hasil yang mencakup penggunaan algoritma C4.5 untuk memprediksi risiko kredit Mcet, serta analisis mengenai kinerja model yang diperoleh. Pengkajian ini mencakup penilaian terhadap data yang digunakan, prosedur pengujian model, serta penekanan pada penerapan prediksi risiko kredit bermasalah dengan menggunakan algoritma C4.5 berbasis model forward selection.

##### 4.1 Hasil Pengolahan Data

Dataset yang digunakan dalam riset ini bersumber dari PT Trust Finance Indonesia dan mencakup total 300 catatan data. Kumpulan data tersebut memiliki 9 atribut, yang terbagi menjadi 5 atribut nominal serta 4 atribut numerik. Adapun metodologi yang diterapkan dalam penelitian ini adalah metode eksperimen dan studi literatur. Rincian karakteristik dataset disajikan pada Tabel 1.

Tabel 1. Deskripsi Karakteristik Dataset PT Trust Finance Indonesia

No	Nama	Type
A1	Nama	Polynomial
A2	Usia	Polynomial
A3	Pekerjaan	Polynomial
A4	Nominal Pinjaman	Polynomial
A5	Status Perkawinan	Binomial
A6	Penadapatan	Polynomial
A7	Periode	Polynomial
A8	Nilai Agunan	Polynomial
A9	Kriteria	Binomial

Penelitian ini menggunakan total 9 atribut, sebagaimana dirinci pada Tabel 1. Atribut-atribut tersebut diklasifikasikan menjadi dua tipe, yaitu 7 atribut bertipe polynomial dan 2 atribut bertipe binomial. Adapun yang termasuk dalam kategori atribut polynomial adalah nama, usia, pekerjaan, jumlah pinjaman, dan penghasilan periode, dan nilai jaminan, sedangkan atribut yang dikategorikan sebagai tipe binomial adalah status perkawinan dan kriteria. Atribut Nama lengkap, Alamat, dan Pendidikan dihilangkan karena tidak relevan untuk menentukan pohon keputusan.

#### 4.2 Hasil Analisa Data

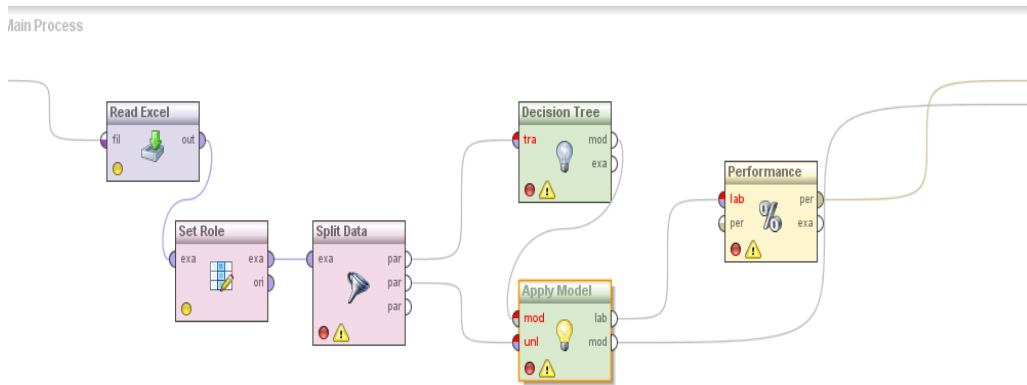
Berikut ini adalah hasil evaluasi data tentang karakteristik yang paling berpengaruh terhadap ketepatan dalam memprediksi risiko kredit macet, berdasarkan studi dan praktik umum dalam pengelolaan risiko kredit. Memilih fitur yang sesuai merupakan langkah penting untuk menciptakan model prediksi yang tepat dan efektif. Seperti terlihat pada gambar 2

Nama Category	Usia Number	Pekerjaan Category	Nominal Kredit Number	Status Category	Pendapatan Number	Periode Number	NilaiJaminan Category	Kriteria Category
Amirullah	41	PNS	6106000	Menikah	132000000	4	151 - 176 %	Lancar
Agustinus	31	PengusahaWir...	7652000	Menikah	54000000	8	176 - 200 %	Lancar
Niekma	48	PengusahaWir...	5016000	Menikah	130000000	5	176 - 200 %	Risiko Mac
Budi	26	Karyawan Swa...	4442000	Menikah	62000000	5	125 - 150 %	Risiko Mac
Davis	55	Pensiunan	5085000	Menikah	43000000	9	151 - 176 %	Lancar
Richard	26	Karyawan Swa...	4924500	Single	140000000	5	125 - 150 %	Risiko Mac
Olla	34	PNS	6191100	Menikah	239000000	5	176 - 200 %	Risiko Mac
Alung	36	TNIPOLRI	5179200	Menikah	109000000	3	151 - 176 %	Lancar

Gambar 2. Analisa Data

#### 4.3 Pengujian Algoritma C4.5 Versi Dasar (Tanpa Optimasi)

Pengujian dan implementasi algoritma C4.5 dalam bentuk dasarnya difokuskan pada penerapan standar algoritma pohon keputusan untuk klasifikasi. Dalam proses ini, tidak ada teknik tambahan yang digunakan untuk augmentasi kinerja, misalnya pemangkasan pohon (pruning). Alur kerja ini secara umum mencakup beberapa tahapan, yang diawali dengan pemrosesan data, dilanjutkan analisis data, dan diakhiri dengan evaluasi hasil. Studi ini secara khusus menguji set data menggunakan C4.5 versi non-optimasi. Perangkat lunak RapidMiner dimanfaatkan untuk mengevaluasi performa dasar algoritma tersebut. Hasil evaluasi ini kemudian ditetapkan sebagai baseline (tolok ukur) untuk komparasi pada penelitian selanjutnya, seperti yang ditunjukkan di gambar 3.



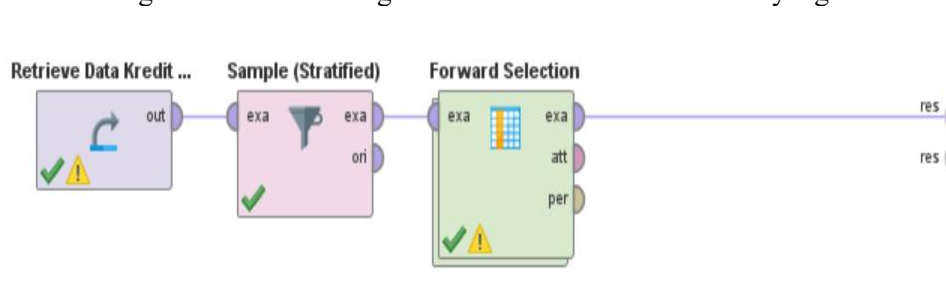
Gambar 3. Implementasi Algoritma C4.5 tidak dengan optimasi

Pada gambar 3. Pengujian ini menunjukkan nilai performa dengan akurasi 83,33%, sensitivitas 96,43%, spesifisitas 43,00%, PPV 85,59%, NPV 69,67%, dan AUC 0,783%. Penelitian sebelumnya yang menyatakan bahwa C4.5 merupakan algoritma efektif untuk memprediksi kelayakan penerimaan beasiswa, sebanding dengan metode manual yang lampau. Walaupun demikian, penerapan C4.5 menjadi kurang optimal ketika digunakan pada sampel data yang lebih besar. Oleh karena itu, optimasi diperlukan untuk melakukan evaluasi terhadap potensi peningkatan akurasi.

#### 4.4 Pengujian hasil Implementasi Algoritma C4.5 Melalui Peningkatan Forward Selection

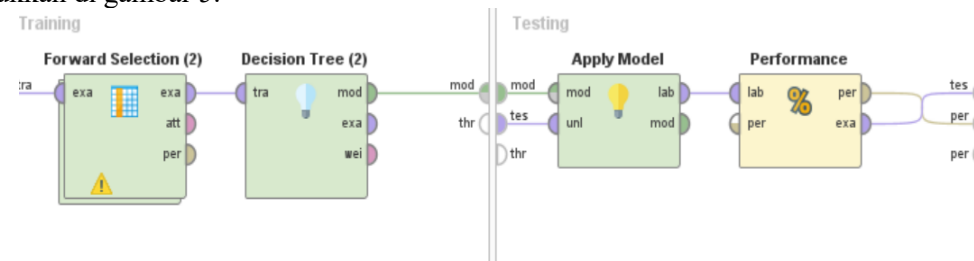
Pada fase ini, sebuah percobaan diterapkan untuk memperbaiki capaian dengan mengadopsi teknik stratified sampling (rujuk Gambar 2). Sebelum proses pemodelan dimulai, data yang sudah terkumpul diolah terlebih dahulu menggunakan teknik pengambilan sampel berstrata.

Sampel data diambil secara random dari keseluruhan populasi, di mana proses ini turut memperhatikan keragaman sebaran data guna membentuk sebuah set data yang baru.



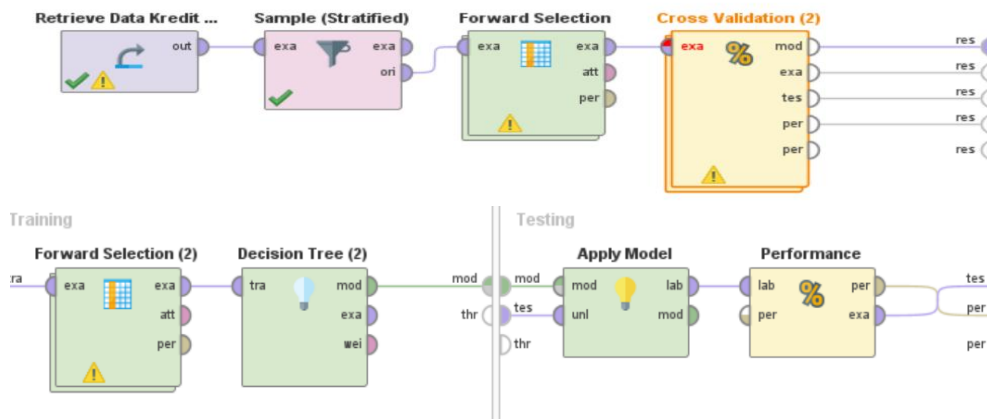
Gambar 4. Implementasi teknik Stratified Sampling menggunakan *RapidMiner*

Keluaran (output) dari sample yang terstratifikasi menunjukkan pemilihan fitur yang paling baik dan juga fitur yang tidak berperan dalam akurasi dengan menggunakan metode Pemilihan forward selection. Proses ini dilakukan secara berkelanjutan. Proses ini akan berhenti ketika fitur yang sedang diuji tidak lagi menunjukkan peningkatan akurasi yang berarti, sebagaimana ditunjukkan di gambar 5.



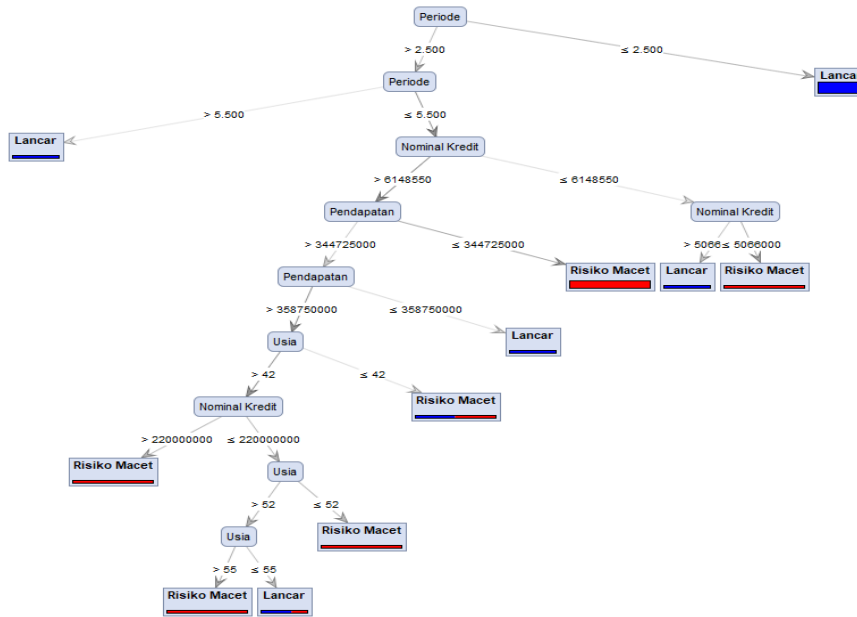
Gambar 5. Metode Forward Selection di Aplikasi *RapidMiner*

Tahapan cross validation setelah melakukan optimasi yang diperoleh dari aplikasi RapidMiner seperti yang terlihat pada gambar 6.



Gambar 6. Metode Forward Selection di Aplikasi *RapidMiner*

Hasil dari pemodelan menunjukkan visualisasi pohon keputusan. Setelah itu, hasil pohon keputusan diuji dengan cross validation untuk mengetahui tingkat akurasi dan AUC, seperti yang terlihat pada gambar 7.



Gambar 7. Tampilan Pohon Keputusan

Tabel 3. Confusion Matrix

	Risiko Kredit	Lancar	Total
Prediksi Risiko Kredit	197	25	224
Prediksi Lancar	9	57	66
Total	206	83	290

Tabel 3 menyajikan Confusion Matrix yang memaparkan hasil dari klasifikasi prediksi. Data ini ditampilkan sebelum dilakukannya perhitungan terhadap nilai akurasi dan AUC. Dalam proses klasifikasi yang dijalankan, dari total 290 sampe, kita memiliki 197 data yang diprediksi sebagai risiko kredit macet dan 25 data sebagai lancar atau layak, sehingga total data yang sebenarnya risiko kredit macet berjumlah 206. Selanjutnya, dari data yang diprediksi risiko kredit macet tetapi sebenarnya Lancar atau layak, ada 9 data yang ditemukan, sedangkan dari data yang diprediksi lancar/layak, terdapat 57 data, sehingga total data yang sebenarnya lancar atau layak mencapai 83. Dengan informasi di atas, kita akan menghitung akurasi, sensitivitas, spesifisitas, PPV, dan NPV menggunakan rumus yang telah ditentukan:

$$Akurasi = \frac{197 + 57}{197 + 75 + 25 + 9} \times 100\% = \frac{254}{290} \times 100\% = 87,59\%$$

$$Sensitivitas = \frac{197}{197 + 9} \times 100\% = \frac{197}{206} \times 100\% = 95,63\%$$

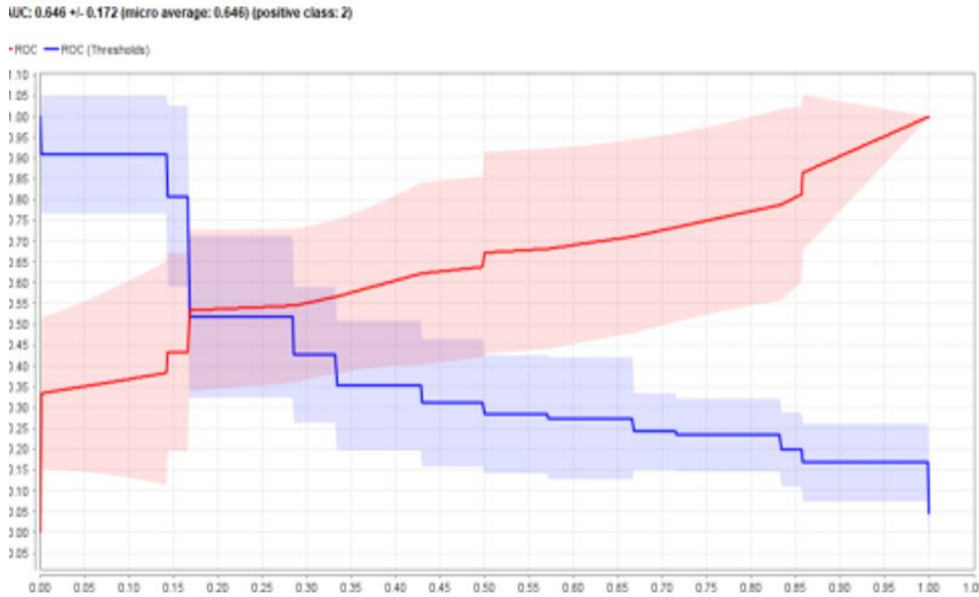
$$Spesifisitas = \frac{25}{25 + 57} \times 100\% = \frac{25}{82} \times 100\% = 30,49\%$$

$$PPV = \frac{57}{197 + 25} \times 100\% = \frac{57}{222} \times 100\% = 88,74\%$$

$$NPV = \frac{57}{57 + 9} \times 100\% = \frac{57}{66} \times 100\% = 86,36\%$$

Kalkulasi yang telah dilakukan menghasilkan beberapa metrik evaluasi. Secara spesifik, nilai yang diperoleh adalah akurasi 87,59%, sensitivitas 95,63%, spesifisitas 30,49%, PPV 88,74%, dan NPV 86,36%. Berdasarkan temuan tersebut, evaluasi dilanjutkan dengan membuat kurva ROC

(Receiver Operating Characteristic) guna menentukan nilai Area Under ROC Curve (AUC). Kurva yang dimaksud disajikan pada Gambar 8.



Gambar 8. Tampilan AUC

Analisis kurva ROC yang disajikan pada Gambar 8 menunjukkan perolehan nilai AUC (Area Under Curve) sebesar 0,646. Lebih lanjut, sebuah kalkulasi dijalankan menggunakan algoritma C4.5 yang tidak melalui proses optimasi, dengan tujuan untuk mengukur seberapa besar peningkatan yang berhasil dicapai. Seluruh temuan dari pengukuran tersebut telah dirangkum dan dapat ditinjau pada Tabel 5 berikut ini:

Tabel 4. Rekapitulasi Pengukuran Sebelum dan Pasca-Perbaikan

Model	Nilai Akurasi	Nilai Sensitivitas	Nilai Spesifisitas	Nilai PPV	Nilai NPV	Nilai AUC
Algoritma C4.5	83,33%	96,43%	43,00%	85,59%	69,67%	0,709%
Algoritma C4.5 dengan Starified Sampling dan Forward Selection	7,59%	95,63%	30,49%	88,74%	86,36%	0,646%

Algoritma C4.5 terbukti mengalami peningkatan akurasi, sebagaimana ditunjukkan oleh hasil pengukuran yang telah dilakukan. Temuan ini sejalan dengan riset sebelumnya yang melaporkan bahwa penerapan metode stratified sampling dan forward selection mampu mendongkrak akurasi algoritma tersebut hingga 4,26%. Studi terdahulu juga mengindikasikan bahwa kombinasi kedua metode itu dapat digunakan untuk mengatasi defisiensi atau kelemahan yang ada pada algoritma C4.5.

**V. KESIMPULAN**

Hasil analisis dan diskusi dalam penelitian ini menghasilkan beberapa kesimpulan yang menjawab rumusan masalah. Pertama, Algoritma C4.5 teridentifikasi mampu melakukan prediksi risiko kredit bermasalah di PT Trust Finance Indonesia dengan capaian akurasi 83.33%. Kedua, ditemukan bahwa akurasi mengalami peningkatan menjadi 87.59% setelah Algoritma C4.5 dioptimasi menggunakan Metode Forward Selection. Poin terakhir menegaskan bahwa implementasi Metode Forward Selection pada optimasi Algoritma C4.5 memberikan kontribusi kenaikan akurasi sebesar 4.26% untuk prediksi kredit macet di PT Trust Finance Indonesia.

## VI. UCAPAN TERIMA KASIH

Dengan rasa hormat penulis mengucapkan terima kasih kepada pihak-pihak yang telah membantu dalam penyusunan Jurnal Ilmiah ini:

1. Bapak Dr. Mochamad Mukti ALI, S.T., M.M. selaku Rektor Universitas Indonesia Membangun Bandung.
2. Ibu Dr. Hj. Erna Herlinawati, S.E., M.Si., selaku Wakil Rektor Universitas Indonesia Membangun Bandung.
3. Bapak Debi Irawan, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Komputer Universitas Indonesia Membangun.
4. Bapak Febriyansyah Ramadhan, S.Kom., M.T. selaku Dosen Pembimbing yang telah memberikan ilmu dan meluangkan waktu untuk membimbing penulis sehingga penulis dapat menyelesaikan laporan magang ini.
5. Bapak Suparman Sulina selaku Direktur Utama PT. Trust Finance Indonesia, Tbk.
6. Bapak Budi Hartono selaku Pembimbing Magang di PT. Trust Finance Indonesia, Tbk.
7. Para Dosen dan Staff Pengajar Program Studi Manajemen Universitas Indonesia Membangun Bandung yang telah mengajar, mengarahkan dan membantu Penulis selama masa perkuliahan.

## VII. REFERENSI

- ADI, S. (2015). PREDIKSI KREDIT MACET MELALUI PERILAKU NASABAH PADA KOPERASI SIMPAN PINJAM DENGAN MENGGUNAKAN METODE ALGORITMA KLASIFIKASI C4.5. *JURNAL DISPROTEK*, 1-13.
- Anton Saputra, T. Y. (2022). OPTIMASI CHI SQUARE DAN PERBAIKAN TEKNIK PRUNNING UNTUK PENINGKATAN AKURASI ALGORITMA C4.5 DALAM MODEL KASUS PREDIKSI KETERLAMBATAN BIAYA KULIAH. *JIKO (Jurnal Informatika dan Komputer)*, 6(2), 231–24.
- Ardiyansyah, R. S. (2023). Peningkatan akurasi metode C4.5 untuk memprediksi kelayakan kredit berbasis stratified sampling dan optimize selection. *Kumpulan jurnaL Ilmu Komputer (KLIK)*, 10(2), 239-248.
- Arie Yandi Saputra, Y. P. (2021). Penerapan metode Technology Acceptance Model terhadap faktor kepercayaan dan risiko dalam penggunaan aplikasi fintech. *Cogito Smart Journal*, 7(2), 290-304.
- Bentar Candra P, K. T. (2024). Optimasi algoritma C4.5 ,menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Mahasiswa Penerima Beasiswa. *Kajian Ilmiah Informatika dan Kompute*, 4(6), 3005-3012.
- Eka Praja Wiyata Mandala, D. E. (2018). Prediksi jumlah pemberian kredit kepada nasabah di Bank Perkreditan Rakyat dengan algoritma C 4.5. *UPI YPTK Jurnal KomTekInfo V*, 5(1), 70-80.
- Ita Kumaratih, E. N. (2017). Pengaruh internal auditor dan pengendalian internal pemberian kredit mikro terhadap kolektibilitas piutang pada PT Bank Mandiri (Persero) Tbk Cabang Bandung Mohammad Toha. *Jurnal Indonesia Membangun. JURNAL INDO*, 16, 100-112.
- Kevin Manurip, D. I. (2022). Analisis Sentimen Distribusi Vaksin COVID-19 di Indonesia Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Ilmiah Universitas Batanghari Jambi*, 22(2), 1205-1208.

- 
- Lusi Asameta, Y. P. (2024). Sistem Pendukung Keputusan Penerimaan Bantuan Sosial Perikanan. *untuk Nelayan menggunakan Metode Simple Additive Weighting (SAW)*, 3(3), 175-188.
- Menarianti, I. (2015). Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi. *Jurnal Ilmiah Teknosains*, 1(1), 36-45.
- Riyandi Nur Sumawidjaya, R. A. (2015). Analisis pengaruh tingkat efisiensi, risiko kredit, risiko likuiditas, rentabilitas, serta permodalan terhadap kinerja keuangan perbankan: Studi kasus pada bank asing Standard Chartered periode 2009–2013. *Jurnal Indonesia Membangun*, 14(1), 78-93.