

Model Data Mining untuk Penetapan Plafon Kredit dengan Algoritma C4.5

Frans Mikael Sinaga¹, Jefri Junifer Pangaribuan^{2*}, Aulia Rizky Muhammad Hendrik Noor Asegaff³, Wenripin Chandra⁴, Riche⁵

^{1,4}Informatics Department, Faculty of AI and Data Sciences, Universitas Pelita Harapan, Medan, Indonesia

²Information Systems Department, School of Information Systems, Universitas Bina Nusantara, Jakarta, Indonesia

³Teknik Informatika, Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin, Indonesia

⁵Sistem Informasi, Fakultas Informatika, Universitas Mikroskil, Medan, Indonesia

Email: ¹frans.sinaga@uph.edu, ^{2*}jefri.pangaribuan@binus.edu, ³aulia.rizky.m.h.n.a@gmail.com,

⁴wenripin@lecturer.uph.edu, ⁵riche@mikroskil.ac.id

Abstract

Manual credit limit determination in distributor companies is often subjective and inconsistent, increasing the risk of bad debts. This research aims to design an objective data mining model to support customer credit limit decisions at CV. XYZ. The method used is the Decision Tree with the C4.5 algorithm, applied to 66 historical records of customer payment data. Data analysis was performed by calculating Entropy and Information Gain values to build the decision tree, which was then validated using RapidMiner Studio software. The research successfully built a valid and consistent classification model. The "Piutang" (receivables/transaction volume per invoice) attribute was identified as the main determinant (root node), followed by the "Pembayaran" (payment history) attribute as a branch node. This model generates three interpretable decision rules, including the discovery of a risky pattern where high-volume customers with poor payment histories are associated with large credit limits. The proposed model can be implemented as a decision support tool to standardize credit policies, reduce subjectivity, and minimize the company's financial risk.

Keywords: Data Mining, Decision Tree, C4.5 Algorithm, Credit Limit, Risk Management.

Abstrak

Penetapan plafon kredit secara manual pada perusahaan distributor seringkali bersifat subjektif dan tidak konsisten, sehingga meningkatkan risiko piutang tak tertagih. Penelitian ini bertujuan untuk merancang sebuah model *data mining* yang objektif untuk mendukung keputusan penetapan plafon kredit pelanggan di CV. XYZ. Metode yang digunakan adalah *Decision Tree* dengan algoritma C4.5, yang diterapkan pada 66 record data historis pelunasan piutang. Analisis data dilakukan dengan menghitung nilai *Entropy* dan *Information Gain* untuk membangun pohon keputusan, yang kemudian divalidasi menggunakan perangkat lunak RapidMiner Studio. Hasil penelitian berhasil membangun model klasifikasi yang valid dan konsisten. Atribut "Piutang" (volume transaksi per faktur) teridentifikasi sebagai faktor penentu utama (simpul akar), diikuti oleh atribut "Pembayaran" (riwayat kelancaran) sebagai simpul cabang. Model ini menghasilkan tiga aturan keputusan yang dapat diinterpretasikan, termasuk mengungkap pola berisiko di mana pelanggan bervolume besar dengan pembayaran buruk terasosiasi dengan plafon besar. Model yang diusulkan dapat diimplementasikan sebagai alat bantu keputusan untuk menstandardisasi kebijakan kredit, mengurangi subjektivitas, dan meminimalkan risiko finansial perusahaan.

Kata Kunci: Data Mining, Decision Tree, Algoritma C4.5, Plafon Kredit, Manajemen Risiko.

1. PENDAHULUAN

Seiring dengan pesatnya perkembangan era digital, lanskap bisnis global terus bertransformasi, memungkinkan transaksi penjualan berlangsung tanpa batasan ruang dan waktu (Harto et al., 2023). Meskipun demikian, model bisnis tradisional, terutama pada sektor distribusi, masih memegang peranan penting dalam rantai pasok ekonomi (Putri & Marsono, 2025). Perusahaan distributor, seperti CV. XYZ yang bergerak di bidang distribusi bahan bangunan, sering kali tidak dapat menerapkan sistem penjualan *cash and carry* sepenuhnya. Ketergantungan pada sistem penjualan kredit menjadi sebuah keniscayaan untuk menjaga loyalitas pelanggan dan volume penjualan. Namun, fasilitas kredit ini membawa risiko yang melekat, terutama risiko kredit macet atau keterlambatan pembayaran. Proses penetapan batas kredit (plafon) yang dilakukan secara manual dan subjektif sering kali menjadi akar permasalahan. Keputusan yang hanya didasarkan pada intuisi atau hubungan personal tanpa analisis data yang mendalam dapat menyebabkan kesalahan dalam menilai kapasitas bayar pelanggan. Akibatnya, perusahaan sering menghadapi piutang yang melebihi jangka waktu yang ditetapkan, yang pada akhirnya mengganggu arus kas dan mengurangi profitabilitas. Kondisi inilah yang menjadi tantangan utama bagi perusahaan, di mana terdapat kebutuhan mendesak untuk beralih dari pengambilan keputusan subjektif ke pendekatan yang lebih sistematis dan berbasis data.

Untuk mengatasi permasalahan tersebut, penelitian ini mengintegrasikan kerangka kerja Sistem Pendukung Keputusan (SPK) atau *Decision Support Systems* (DSS) dengan teknik *data mining* (Indriyani et al., 2024). Dalam konteks manajemen risiko keuangan, SPK berbasis data bertujuan untuk mentransformasi data historis—yang seringkali terabaikan—menjadi wawasan yang dapat ditindaklanjuti (*actionable insights*), menggantikan proses manual yang rentan terhadap bias kognitif dan inkonsistensi (Han et al., 2012). Metode *Decision Tree* (Pohon Keputusan) dipilih sebagai motor analitik utama untuk SPK ini. Alasan pemilihannya tidak hanya terbatas pada kemampuan prediktifnya, tetapi juga karena keunggulan utamanya dalam *interpretability* (Barus et al., 2024). Algoritma C4.5, sebagai varian *Decision Tree*, secara spesifik digunakan karena kemampuannya yang efisien dalam menangani data atribut numerik dan kategorikal secara bersamaan (Pangaribuan & Putra, 2024). Berbeda dengan model *black box* (seperti SVM atau Jaringan Saraf Tiruan), C4.5 menghasilkan model *white box* berupa aturan 'JIKA-MAKA' yang eksplisit. Transparansi ini krusial dalam konteks bisnis, karena memungkinkan manajer non-teknis untuk memahami, memvalidasi, dan pada akhirnya memercayai dasar pengambilan keputusan yang direkomendasikan oleh model.

Untuk memetakan posisi dan kebaruan (*novelty*) penelitian ini, dilakukan tinjauan literatur terstruktur terhadap penelitian-penelitian sebelumnya. Sebagian besar penelitian yang ada (Ibrahim & Khikmah, 2024; Sibarani et al., 2023) menerapkan C4.5 pada lembaga keuangan seperti koperasi dan bank dengan fokus pada masalah klasifikasi biner: menentukan kelayakan kredit (Layak/Tidak Layak). Beberapa penelitian lain, seperti (Saputra & Fatah, 2025), melangkah lebih jauh dengan melakukan segmentasi profil risiko (Rendah, Sedang, Tinggi), namun tidak sampai pada rekomendasi plafon spesifik. Studi lain yang membandingkan metode, seperti (Eldo et al., 2024) dengan SVM, mungkin mencapai akurasi tinggi namun menghasilkan model *black box* yang sulit diinterpretasi, sehingga kurang ideal untuk SPK yang membutuhkan transparansi manajerial. Dari tinjauan ini, terlihat jelas adanya dua kesenjangan penelitian (*research gap*) yang signifikan. Pertama, fokus utama penelitian sebelumnya adalah pada penentuan kelayakan, bukan pada pertanyaan praktis "jika layak, berapa nilai plafon yang harus

diberikan?". Kedua, konteks penelitian didominasi oleh sektor B2C (perbankan/koperasi) yang datanya (seperti gaji, tanggungan) berbeda dengan sektor B2B (distributor). Penelitian ini mengisi kedua celah tersebut dengan menerapkan C4.5 untuk penetapan kategori plafon (Kecil/Besar) dalam konteks B2B yang unik, di mana variabel transaksional (seperti volume piutang per faktur dan riwayat kelancaran pembayaran) menjadi prediktor yang lebih relevan.

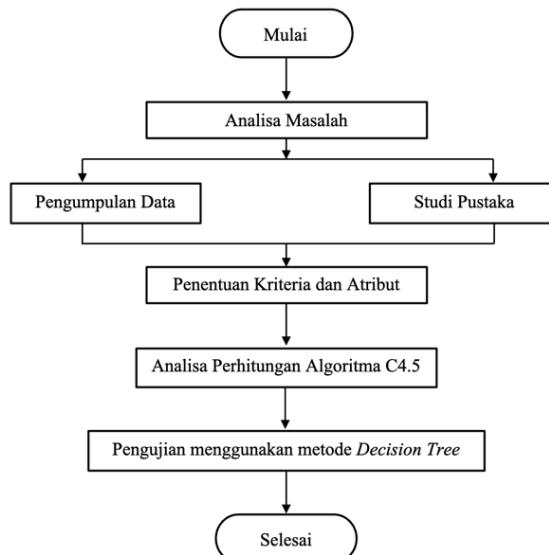
Berdasarkan latar belakang, integrasi kerangka SPK, dan analisis kesenjangan tersebut, tujuan utama dari penelitian ini adalah untuk merancang dan mengimplementasikan model *data mining* menggunakan metode *Decision Tree* dengan algoritma C4.5 untuk menentukan plafon kredit pelanggan di CV. XYZ. Secara spesifik, penelitian ini bertujuan untuk (1) menerapkan algoritma C4.5 pada data historis piutang pelanggan untuk membentuk pohon keputusan yang transparan; (2) mengidentifikasi atribut-atribut transaksional yang paling signifikan dalam menentukan kapasitas pembayaran pelanggan; dan (3) menghasilkan serangkaian aturan (*rules*) yang dapat diinterpretasikan dan diintegrasikan ke dalam SPK untuk proses pengambilan keputusan. Harapan dari penelitian ini adalah model yang dihasilkan dapat menjadi sebuah sistem pendukung keputusan yang akurat dan objektif, membantu manajemen dalam mengurangi risiko piutang tak tertagih, serta meningkatkan konsistensi dan efisiensi dalam proses pemberian kredit kepada pelanggan.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik *data mining* untuk membangun model klasifikasi. Metode yang digunakan adalah *Decision Tree* dengan algoritma C4.5 untuk menganalisis data historis piutang pelanggan dan menentukan plafon kredit yang sesuai.

2.1 Tahapan Penelitian

Proses penelitian ini dirancang secara sistematis yang terdiri dari beberapa tahapan utama, mulai dari identifikasi masalah hingga analisis hasil akhir. Setiap tahapan dilakukan secara berurutan untuk memastikan penelitian berjalan sesuai dengan tujuan yang telah ditetapkan. Alur tahapan penelitian diilustrasikan pada Gambar 1.



Gambar 1. Alur tahapan penelitian

Berikut adalah penjelasan rinci dari setiap tahapan yang dilakukan dalam penelitian ini:

1. Analisa Masalah: Tahap awal ini berfokus pada identifikasi dan pemahaman mendalam terhadap masalah yang dihadapi oleh CV. XYZ, yaitu kesulitan dalam menentukan plafon kredit pelanggan secara objektif yang berakibat pada tingginya risiko kredit macet. Analisis ini menjadi dasar untuk merumuskan tujuan dan ruang lingkup penelitian.
2. Pengumpulan Data: Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh langsung dari perusahaan. Pengumpulan data dilakukan melalui beberapa metode:
 - a. Observasi: Melakukan pengamatan dan pengumpulan data laporan piutang pelanggan selama periode tahun 2020. Total data yang berhasil dikumpulkan adalah sebanyak 66 *record* transaksi pelunasan.
 - b. Wawancara: Melakukan diskusi dengan pihak manajer perusahaan untuk memahami proses bisnis yang berjalan, kebijakan pemberian kredit saat ini, dan variabel-variabel yang dianggap relevan dalam penilaian pelanggan.
3. Studi Pustaka: Tahap ini dilakukan bersamaan dengan pengumpulan data untuk memperkuat landasan teori. Studi pustaka mencakup penelusuran jurnal ilmiah, buku, dan artikel yang relevan dengan *data mining*, metode *Decision Tree*, algoritma C4.5, dan manajemen risiko kredit.
4. Penentuan Kriteria dan Atribut (*Data Preparation*): Data mentah yang telah dikumpulkan kemudian diproses pada tahap ini. Proses ini meliputi pembersihan data (*data cleaning*) dan transformasi data (*data transformation*) dari bentuk numerik menjadi kategorikal agar sesuai dengan format yang dibutuhkan oleh algoritma C4.5. Sebanyak lima atribut ditentukan, di mana empat atribut menjadi prediktor dan satu atribut menjadi target kelas (label). Kriteria untuk setiap atribut didefinisikan secara spesifik seperti yang dijelaskan pada Tabel 1.

Tabel 1. Kriteria penilaian atribut

Kriteria	Syarat	Klasifikasi
Jenis Pembayaran	Pembayaran <= 35 Hari Pembayaran > 35 Hari	Tunai Kredit
Piutang	Transaksi <= 5 Juta / Faktur Transaksi > 5 Juta / Faktur	Sedikit Banyak
Transaksi	Jumlah Transaksi < 2 Faktur Jumlah Transaksi >= 2 Faktur	Sedikit Banyak
Pembayaran	Jangka Waktu Pembayaran <= 75 Hari Jangka Waktu Pembayaran > 75 Hari	Sangat Bagus Tidak Bagus
Plafon	Jumlah Pelunasan <= Rp. 10.000.000,- / Faktur Jumlah Pelunasan > Rp. 10.000.000,- / Faktur	Kecil Besar

5. Analisis Perhitungan Algoritma C4.5: Tahap inti dari penelitian ini adalah penerapan algoritma C4.5 pada *dataset* yang telah disiapkan. Proses ini diawali dengan perhitungan nilai *Entropy* untuk mengukur tingkat heterogenitas data pada kelas target. Selanjutnya, dihitung nilai *Information Gain* untuk setiap atribut prediktor. Atribut dengan nilai Gain tertinggi dipilih sebagai simpul akar (*root node*). Proses ini diulang secara rekursif untuk setiap cabang hingga semua data dalam satu cabang memiliki kelas yang sama atau tidak ada atribut tersisa.
6. Pengujian dan Analisis Hasil: Pohon keputusan yang dihasilkan dari perhitungan manual kemudian diuji dan divalidasi dengan menggunakan perangkat lunak *data mining* RapidMiner Studio. Hasil dari kedua metode (manual dan *software*)

dibandingkan untuk memastikan konsistensi dan akurasi model. Hasil akhir berupa pohon keputusan dan serangkaian aturan (*rules*) dianalisis untuk ditarik kesimpulan yang dapat memberikan wawasan bagi perusahaan dalam menentukan plafon kredit.

2.2 Dataset dan Transformasi Data

Tahap ini krusial untuk memastikan kualitas data yang digunakan dalam pemodelan. Proses ini mencakup analisis deskriptif *dataset* mentah dan justifikasi transparansi untuk proses transformasi data.

1. Analisis Deskriptif *Dataset*

Data yang digunakan adalah 66 *record* laporan pelunasan piutang tahun 2020. *Dataset* ini memiliki 4 atribut prediktor dan 1 atribut kelas (target). Distribusi frekuensi dari *dataset* yang telah dikategorisasi disajikan pada Tabel 2. Distribusi ini menyoroti ketidakseimbangan data (*imbalanced dataset*), terutama pada kelas target (Plafon), di mana kelas "Kecil" (62 kasus) jauh lebih dominan daripada kelas "Besar" (4 kasus). Ketidakseimbangan ini juga terlihat pada atribut prediktor lainnya, yang mencerminkan kondisi bisnis aktual perusahaan.

Tabel 2. Distribusi Frekuensi Atribut *Dataset* (N=66)

Atribut	Kategori	Jumlah (Frekuensi)	Persentase
Plafon (Target)	Kecil	62	93.9%
	Besar	4	6.1%
Piutang	Sedikit	54	81.8%
	Banyak	12	18.2%
Pembayaran	Sangat Bagus	44	66.7%
	Tidak Bagus	22	33.3%
Transaksi	Sedikit	58	87.9%
	Banyak	8	12.1%
Jenis Pembayaran	Kredit	59	89.4%
	Tunai	7	10.6%

2. Transformasi dan Diskretisasi Data

Data mentah yang bersifat numerik (seperti jumlah faktur dan lama hari pembayaran) perlu ditransformasi menjadi data kategorikal (diskretisasi) agar dapat diolah oleh algoritma C4.5. Proses ini tidak dilakukan secara arbitrer, melainkan didasarkan pada kebijakan bisnis dan observasi operasional di CV. XYZ, seperti yang dirangkum pada Tabel 3.

Tabel 3. Kriteria dan Justifikasi Transformasi Atribut

Kriteria	Syarat (Data Mentah)	Klasifikasi	Justifikasi Ambang Batas
Plafon (Label)	Jumlah Pelunasan <= 10 Juta	Kecil	Ambang batas internal perusahaan untuk membedakan antara plafon standar (Kecil) dan plafon risiko tinggi (Besar).
	Jumlah Pelunasan > 10 Juta	Besar	
Pembayaran	Jangka Waktu <= 75 Hari	Sangat Bagus	Batas waktu maksimal (SOP) yang diberikan perusahaan kepada pelanggan untuk pelunasan (sesuai hasil wawancara). Pelunasan dalam rentang ini dianggap disiplin.
	Jangka Waktu > 75 Hari	Tidak Bagus	
Piutang	Transaksi < 5 Juta / Faktur	Sedikit	Nilai 5 Juta adalah plafon kredit maksimal yang biasa diberikan untuk pelanggan baru (sesuai hasil wawancara). Ini menjadi ambang batas logis untuk transaksi "Standar" vs "Besar".
	Transaksi >= 5 Juta / Faktur	Banyak	

Transaksi	Jumlah Transaksi < 2 Faktur	Sedikit	Ambang batas bisnis untuk membedakan antara pelanggan <i>one-time</i> / infrekuen (Sedikit) dan pelanggan <i>repeat order</i> / reguler (Banyak).
	Jumlah Transaksi >= 2 Faktur	Banyak	
Jenis Pembayaran	Pembayaran <= 35 Hari	Tunai	Definisi operasional perusahaan untuk pembayaran yang dianggap "Cepat" atau setara tunai (misal, kebijakan Net 30 + 5 hari), dibandingkan pembayaran jangka panjang (Kredit).
	Pembayaran > 35 Hari	Kredit	

2.3 Metode *Decision Tree* (Algoritma C4.5)

Decision Tree adalah salah satu metode klasifikasi dalam *data mining* yang memodelkan serangkaian keputusan dalam struktur hirarkis menyerupai pohon. Setiap simpul internal (*internal node*) pada pohon merepresentasikan sebuah tes terhadap suatu atribut, setiap cabang (*branch*) merepresentasikan hasil dari tes tersebut, dan setiap simpul daun (*leaf node*) merepresentasikan label kelas atau keputusan akhir (Resti et al., 2023).

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang digunakan untuk membangun *Decision Tree*. Proses pembentukan pohon dimulai dengan memilih atribut terbaik sebagai simpul akar (*root node*). Pemilihan ini didasarkan pada konsep Information Gain tertinggi, yang dihitung dari nilai *Entropy* (Rusyana et al., 2023).

1. *Entropy*

Entropy adalah ukuran ketidakpastian atau keragaman (*impurity*) dalam sebuah himpunan data. Nilai *entropy* berkisar antara 0 dan 1. Nilai 0 menunjukkan bahwa semua data dalam himpunan tersebut termasuk dalam satu kelas yang sama (homogen), sedangkan nilai 1 menunjukkan bahwa data terdistribusi secara merata di antara semua kelas (heterogen maksimum). Rumus untuk menghitung *entropy* adalah sebagai berikut:

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

Dimana:

S = Himpunan kasus (dataset)

n = Jumlah kelas pada atribut target

p_i = Proporsi jumlah kasus pada kelas i terhadap total kasus dalam himpunan S

2. *Information Gain*

Information Gain digunakan untuk mengukur seberapa efektif sebuah atribut dalam memisahkan data ke dalam kelas-kelas yang lebih murni (homogen). Atribut dengan nilai Gain tertinggi akan memberikan pemisahan data terbaik. *Information Gain* dihitung dengan cara mengurangi *entropy* awal (sebelum pemisahan) dengan rata-rata *entropy* setelah data dipisah berdasarkan atribut tersebut. Rumus *Information Gain* adalah sebagai berikut:

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

Di mana:

S = Himpunan kasus (dataset)

A = Atribut yang diuji

$\text{Values}(A)$ = Himpunan semua nilai yang mungkin untuk atribut A

$|S_v|$ = Jumlah kasus untuk nilai v dari atribut A

$|S|$ = Jumlah total kasus dalam himpunan S

Algoritma akan memilih atribut dengan nilai $Gain(S, A)$ tertinggi sebagai simpul pemisah. Proses ini kemudian diulangi untuk setiap sub-himpunan data yang terbentuk hingga semua kasus dalam sub-himpunan tersebut memiliki kelas yang sama atau tidak ada lagi atribut yang dapat digunakan untuk memisah data lebih lanjut.

2.4 Perangkat Lunak Penelitian

Untuk memastikan validitas dan efisiensi dalam proses analisis data, penelitian ini menggunakan perangkat lunak RapidMiner Studio. RapidMiner adalah platform *data science* yang bersifat *open-source* dan menyediakan lingkungan visual untuk proses *data mining*, *machine learning*, dan analisis prediktif (Sudarsono et al., 2021). Dalam penelitian ini, RapidMiner Studio digunakan untuk beberapa tujuan utama: (1) Mengimpor dan mempersiapkan *dataset* laporan piutang pelanggan; (2) Menerapkan operator *Decision Tree* untuk secara otomatis membangun model klasifikasi berdasarkan algoritma C4.5; (3) Memvisualisasikan hasil akhir berupa pohon keputusan dalam bentuk grafis dan deskripsi aturan (*rules*). Penggunaan perangkat lunak ini bertujuan untuk memvalidasi hasil perhitungan manual serta mendemonstrasikan bagaimana proses analisis dapat diotomatisasi dalam skenario bisnis nyata, sehingga memberikan hasil yang lebih cepat dan konsisten, terutama saat berhadapan dengan volume data yang besar (Nofitri & Irawati, 2019).

2.5 Evaluasi Kinerja Model

Sesuai dengan permintaan peninjau, evaluasi kinerja model tidak hanya didasarkan pada validasi struktur pohon, tetapi juga menggunakan metrik kuantitatif standar untuk model klasifikasi. Mengingat ukuran *dataset* yang terbatas ($N=66$), model dievaluasi kinerjanya terhadap data pelatihan itu sendiri (*resubstitution evaluation*). Meskipun metode ini cenderung memberikan hasil yang optimis, ini adalah langkah penting untuk mengukur seberapa baik model fit atau mencocoki data historis yang ada. Evaluasi dilakukan menggunakan *Confusion Matrix*.

1. Confusion Matrix

Confusion Matrix (Matriks Kebingungan) adalah tabel yang merangkum kinerja model klasifikasi. Tabel ini membandingkan kelas aktual dengan kelas yang diprediksi oleh model. Untuk masalah biner seperti ini, matriks akan berisi empat nilai: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dalam konteks ini, kelas "Besar" dianggap sebagai Positif dan "Kecil" sebagai Negatif.

2. Metrik Evaluasi

Berdasarkan nilai dari Confusion Matrix, metrik-metrik berikut dihitung untuk menilai kinerja model:

- a. Akurasi (*Accuracy*): Mengukur proporsi total prediksi yang benar (TP+TN) dari keseluruhan data.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

- b. Presisi (*Precision*): Mengukur proporsi kasus positif yang diprediksi dengan benar (TP) dari total kasus yang diprediksi sebagai positif (TP+FP). Metrik ini penting untuk mengetahui seberapa "bersih" prediksi positif; dalam kasus ini, "Dari semua yang diprediksi Plafon Besar, berapa yang sebenarnya benar-benar Besar?".

$$Presisi = \frac{TP}{TP + FP}$$

- c. Sensitivitas (*Recall*): Mengukur proporsi kasus positif aktual (TP) yang berhasil diidentifikasi dengan benar oleh model (TP+FN). Metrik ini penting untuk mengetahui seberapa baik model menemukan semua kasus positif; dalam kasus ini,

"Dari semua pelanggan Plafon Besar yang sebenarnya, berapa yang berhasil ditemukan model?".

$$Sensitivitas (Recall) = \frac{TP}{TP + FN}$$

- d. *F1-Score*: Rata-rata harmonik dari *Presisi* dan *Recall*, memberikan satu nilai yang menyeimbangkan kedua metrik tersebut, terutama berguna pada *dataset* yang tidak seimbang.

$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall}$$

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari penerapan algoritma C4.5 pada dataset piutang pelanggan CV. XYZ. Proses analisis diawali dengan perhitungan manual, dilanjutkan dengan validasi dan evaluasi kinerja menggunakan RapidMiner, dan diakhiri dengan pembahasan mendalam mengenai implikasi dari aturan keputusan yang dihasilkan.

3.1 Pembangunan dan Analisis Pohon Keputusan

Proses pembentukan pohon keputusan dimulai dengan menghitung nilai *Entropy* total dari *dataset* dan nilai *Information Gain* untuk setiap atribut guna menentukan simpul akar (*root node*).

1. Penentuan Simpul Akar (Node 1)

Proses pembentukan pohon keputusan dimulai dengan menghitung nilai *Entropy* total dari *dataset* ($N=66$) yang memiliki Entropy(Total) = 0.3298.

$$\begin{aligned} Entropy(Total) &= (-4/66 \times \log_2(4/66)) + (-62/66 \times \log_2(62/66)) \\ &= 0.3298 \end{aligned}$$

Nilai ini relatif rendah, mencerminkan ketidakseimbangan data (Tabel 1) di mana kelas "Kecil" sangat dominan (93.9%). Selanjutnya, *Information Gain* (IG) dihitung untuk setiap atribut (Tabel 4).

Tabel 4. Perhitungan *Gain* untuk Penentuan Simpul Akar

Atribut	Nilai Atribut	Jlh Kasus	Besar	Kecil	Entrophy	Gain
Total		66	4	62	0.3298	
Jenis Pembayaran	Tunai	7	1	6	0.5917	0.0078
	Kredit	59	3	56	0.2899	
Piutang	Banyak	12	4	8	0.9183	0.1628
	Sedikit	54	0	54	0	
Transaksi	Banyak	8	1	7	0.5435	0.0058
	Sedikit	58	3	55	0.2936	
Pembayaran	Sangat Bagus	44	1	43	0.1564	0.0339
	Tidak Bagus	22	3	19	0.5746	

Analisis Tabel 4 menunjukkan bahwa atribut "Piutang" ($IG = 0.1628$) terpilih sebagai simpul akar. Secara bisnis, ini adalah temuan yang sangat signifikan. Nilai IG "Piutang" jauh lebih tinggi daripada atribut lainnya (misal, "Pembayaran" hanya

0.0339). Ini mengindikasikan bahwa pertanyaan pertama dan terpenting yang harus diajukan dalam menentukan plafon kredit di CV. XYZ adalah "Berapa besar volume transaksi pelanggan per faktur?". Atribut ini memiliki daya pembeda (*discriminatory power*) tertinggi. Logika bisnisnya adalah, volume transaksi adalah proksi utama untuk skala kebutuhan pelanggan; pelanggan yang membutuhkan plafon "Besar" pastilah mereka yang melakukan transaksi "Banyak". Model ini mengonfirmasi bahwa cabang "Piutang = Sedikit" memiliki Entropy 0 (54 kasus, 0 "Besar"), yang berarti semua pelanggan dengan transaksi kecil secara eksklusif jatuh ke plafon "Kecil". Ini adalah aturan yang sangat murni dan kuat, yang secara instan menyederhanakan 82% (54 dari 66) kasus.

2. Penentuan Simpul Cabang (Node 1.1)

Perhitungan dilanjutkan hanya untuk sub-himpunan data yang "rumit", yaitu 12 kasus di mana Piutang = "Banyak". Node ini memiliki Entropy yang tinggi (0.9183), menunjukkan ketidakpastian besar (4 kasus "Besar", 8 kasus "Kecil"). Perhitungan IG (Tabel 5) dilakukan untuk menentukan atribut terbaik guna memecah ketidakpastian ini.

Tabel 5. Perhitungan *Gain* untuk Node Piutang = "Banyak"

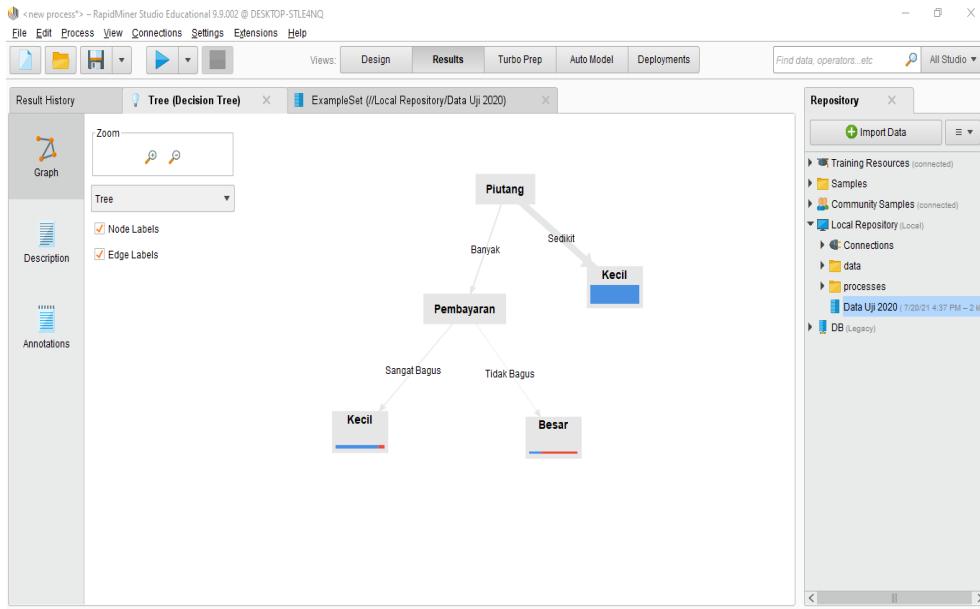
Atribut	Nilai Atribut	Jlh Kasus	Besar	Kecil	Entropy	Gain
Total (Piutang=Banyak)		12	4	8	0.9183	
Jenis Pembayaran	Tunai	2	1	1	1.0000	0.0172
	Kredit	10	7	3	0.8812	
Transaksi	Banyak	3	1	2	0.9183	0.0000
	Sedikit	9	3	6	0.9183	
Pembayaran	Sangat Bagus	8	1	7	0.5435	0.2854
	Tidak Bagus	4	3	1	0.8112	

Atribut "Pembayaran" ($IG = 0.2854$) terpilih sebagai simpul cabang. Ini kembali menunjukkan logika bisnis yang kuat: setelah diketahui pelanggan melakukan transaksi besar, pertanyaan terpenting berikutnya adalah "Bagaimana riwayat pembayaran mereka?". Atribut ini jauh lebih unggul daripada "Jenis Pembayaran" ($IG=0.0172$) dan "Transaksi" ($IG=0.0000$), yang ternyata tidak relevan untuk grup ini. Urutan simpul (Piutang -> Pembayaran) ini mencerminkan proses berpikir analitis yang ideal: pertama, nilai transaksi, kemudian, riwayat kedisiplinan. Kedua cabang dari simpul "Pembayaran" ini langsung menjadi simpul daun, mengakhiri pembentukan pohon.

3.2 Evaluasi dan Validasi Model

1. Validasi Struktur Pohon

Model yang dihasilkan secara manual divalidasi menggunakan RapidMiner Studio. Hasil visualisasi RapidMiner (Gambar 2) mengonfirmasi struktur pohon yang identik. Struktur ini (Piutang sebagai akar, diikuti Pembayaran) divalidasi sebagai representasi matematis yang paling efisien untuk memodelkan data historis perusahaan.



Gambar 2. Hasil pohon keputusan dari RapidMiner

2. Evaluasi Kinerja Klasifikasi

Seperti yang dijelaskan dalam metodologi (sub-bab 2.5), evaluasi kinerja kuantitatif dilakukan pada data pelatihan ($N=66$) untuk mengukur *goodness-of-fit* model. Aturan keputusan yang dihasilkan dari pohon mengklasifikasikan 66 kasus, yang hasilnya dirangkum dalam *Confusion Matrix* pada Tabel 6.

Tabel 6. *Confusion Matrix* (Evaluasi Data Pelatihan)

	Besar (Positif)	Kecil (Negatif)
Besar (Positif)	TP = 3	FN = 1
Kecil (Negatif)	FP = 1	TN = 61

Berdasarkan Tabel 5, model membuat dua kesalahan klasifikasi pada data pelatihan: satu kasus "Besar" diprediksi sebagai "Kecil" (*False Negative*), dan satu kasus "Kecil" diprediksi sebagai "Besar" (*False Positive*). Kinerja model disajikan pada Tabel 7.

Tabel 7. Hasil Metrik Evaluasi Kinerja Model

Metrik	Kelas: "Besar" (Positif)	Kelas: "Kecil" (Negatif)	Total (Rata-rata)
Presisi	$\frac{3}{3+1} = 75.00\%$	$\frac{61}{61+1} = 98.39\%$	
Recall	$\frac{3}{3+1} = 75.00\%$	$\frac{61}{61+1} = 98.39\%$	
F1-Score	$2 \times \frac{0.75 \times 0.75}{0.75 + 0.75} = 75.00\%$	$2 \times \frac{0.9839 \times 0.9839}{0.9839 + 0.9839} = 98.39\%$	
Akurasi			$\frac{3+61}{66} = 96.97\%$

Hasil evaluasi menunjukkan akurasi keseluruhan yang sangat tinggi (96.97%), yang mengindikasikan model C4.5 sangat cocok (fit) dengan data historis. Namun, kinerja pada kelas "Besar" (Presisi/Recall 75%) menunjukkan bahwa model sedikit kesulitan dalam mengidentifikasi kelas minoritas ini dengan sempurna—hal yang wajar terjadi pada dataset yang sangat tidak seimbang. Meskipun demikian, kinerja pada kelas "Kecil" (98.39%) hampir sempurna, dan akurasi keseluruhan tetap sangat kuat.

3.3 Implikasi Aturan Keputusan (*Rules*) dan Pembahasan Kritis

Model *Decision Tree* yang telah dibangun dan dievaluasi menghasilkan tiga aturan (*rules*) final. Pembahasan berikut menganalisis implikasi bisnis dari setiap aturan.

1. Aturan 1: JIKA Piutang = "Sedikit" MAKA Plafon = "Kecil" (Prediksi: Plafon "Kecil" | Akurasi: 54/54 = 100%)

Aturan ini mencakup 82% dari seluruh data. Ini adalah aturan dasar manajemen risiko: pelanggan dengan volume transaksi kecil (< 5 Juta) seharusnya hanya menerima plafon standar (Kecil, < 10 Juta). Akurasinya yang 100% menunjukkan bahwa kebijakan ini sudah berjalan konsisten dan divalidasi oleh data sebagai praktik terbaik yang berisiko rendah.

2. Aturan 2: JIKA Piutang = "Banyak" DAN Pembayaran = "Sangat Bagus" MAKA Plafon = "Kecil" (Prediksi: Plafon "Kecil" | Akurasi: 7/8 = 87.5%)

Aturan ini memberikan wawasan strategis. Secara intuitif, pelanggan bervolume besar dan disiplin mungkin dianggap layak mendapat plafon besar. Namun, data historis menunjukkan bahwa pola yang sukses (di mana pembayaran tetap "Sangat Bagus") justru terjadi ketika plafon mereka tetap dijaga "Kecil" (terkendali). Model ini mengungkap bahwa strategi konservatif ini efektif dan harus dilanjutkan. Kesalahan klasifikasi 1 dari 8 kasus (FN) menunjukkan ada satu pelanggan "Besar" yang cocok dengan profil ini, namun mayoritas (7 dari 8) tetap "Kecil".

3. Aturan 3: JIKA Piutang = "Banyak" DAN Pembayaran = "Tidak Bagus" MAKA Plafon = "Besar" (Prediksi: Plafon "Besar" | Akurasi: 3/4 = 75%)

Ini adalah aturan paling kritis yang diidentifikasi oleh model, yang secara langsung menyoroti kelemahan fatal dalam proses pengambilan keputusan manual sebelumnya. Aturan ini mengungkap adanya pola di mana pelanggan yang paling berisiko (transaksi besar dan riwayat pembayaran buruk) justru terasosiasi dengan plafon kredit terbesar. Ini adalah anomali berbahaya yang mencerminkan logika bisnis yang keliru: manajemen kemungkinan besar terfokus pada "volume penjualan" yang tinggi dan mengabaikan sinyal bahaya "pembayaran macet", sehingga terus memberikan plafon yang lebih besar kepada pelanggan yang sebenarnya tidak layak.

Implikasi dari Aturan 3 ini sangat signifikan terhadap tujuan penelitian untuk meminimalkan risiko finansial. Tiga kasus "*True Positive*" (TP) yang diidentifikasi oleh aturan ini (dan satu *False Positive*) mewakili segmen pelanggan yang paling mungkin menyebabkan kerugian piutang tak tertagih. Dengan mengimplementasikan SPK berbasis model ini, manajemen akan mendapatkan peringatan dini yang kuat. Ketika seorang pelanggan baru atau lama cocok dengan profil (Piutang="Banyak", Pembayaran="Tidak Bagus"), sistem tidak akan merekomendasikan plafon "Besar". Tindakan ini secara langsung memitigasi risiko, mencegah terulangnya kesalahan yang sama, dan menghindari kerugian finansial di masa depan. Aturan ini adalah justifikasi utama untuk adopsi model, karena ia mengubah data historis yang buruk menjadi kebijakan preventif yang kuat.

4. KESIMPULAN

Penelitian ini telah berhasil menjawab permasalahan mengenai penetapan plafon kredit pelanggan di CV. XYZ dengan menerapkan metode *Decision Tree* algoritma C4.5. Model klasifikasi yang dibangun dari 66 *record* data piutang pelanggan secara efektif mentransformasi proses pengambilan keputusan yang sebelumnya subjektif menjadi

objektif dan berbasis data. Temuan kunci dari penelitian ini adalah identifikasi dua atribut paling berpengaruh, yaitu "Piutang" (volume transaksi per faktur) sebagai faktor penentu utama, dan "Pembayaran" (riwayat kelancaran pembayaran) sebagai faktor penentu sekunder untuk pelanggan dengan volume transaksi besar. Konsistensi hasil antara perhitungan manual dan validasi menggunakan perangkat lunak RapidMiner Studio mengonfirmasi validitas model yang dihasilkan.

Implikasi praktis dari penelitian ini adalah tersedianya tiga aturan keputusan yang jelas dan mudah diinterpretasikan. Aturan-aturan ini dapat diimplementasikan secara langsung oleh manajemen CV. XYZ sebagai Sistem Pendukung Keputusan (SPK) untuk menetapkan plafon kredit. Secara khusus, model ini menyoroti segmen pelanggan paling berisiko (Piutang "Banyak" dan Pembayaran "Tidak Bagus") yang terasosiasi dengan plafon "Besar", sebuah anomali yang mencerminkan risiko dari kebijakan manual sebelumnya. Dengan mengadopsi model ini, perusahaan dapat secara signifikan mengurangi risiko piutang tak tertagih, menstandardisasi kebijakan kredit, dan meningkatkan kesehatan arus kas. Untuk penelitian selanjutnya, disarankan untuk menggunakan volume data yang lebih besar dan mencakup periode waktu yang lebih panjang guna meningkatkan akurasi dan menemukan pola yang lebih granular. Selain itu, penelitian di masa depan dapat mengeksplorasi penambahan atribut lain, seperti lama berlangganan atau lokasi pelanggan, serta membandingkan kinerja algoritma C4.5 dengan metode klasifikasi lain untuk menemukan model yang paling optimal.

REFERENCES

- Barus, O. P., Phan, N., Widjaja, A. E., Pangaribuan, J. J., & Romindo, R. (2024). Heart Disease Classification Using Decision Trees. *2024 3rd International Conference on Creative Communication and Innovative Technology, ICCIT 2024*. <https://doi.org/10.1109/ICCIT62134.2024.10701238>
- Eldo, H., Ayuliana, A., Suryadi, D., Chrisnawati, G., & Judijanto, L. (2024). Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan pada Transaksi Online. *Jurnal Minfo Polgan*, 13(2), 1627–1632. <https://doi.org/10.33395/jmp.v13i2.14186>
- Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In J. Han, M. Kamber, & J. Pei (Eds.), *Data Mining (Third Edition)* (Third Edition). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Harto, B., Rukmana, A. Y., Subekti, R., Tahir, R., Waty, E., Situru, A. C., & Sepriano. (2023). *Transformasi Bisnis di Era Digital: Teknologi Informasi dalam Mendukung Transformasi Bisnis di Era Digital* (Efitra, Ed.; Cetakan Pertama). Sonpedia Publishing Indonesia.
- Ibrahim, N. H., & Khikmah, L. (2024). Perbandingan Metode Algoritma C4.5, Naïve Bayes, dan Logistic Regression untuk Penentuan Kelayakan Penerima Kredit. *Teknologi*, 14(2), 85–93. <https://doi.org/10.26594/teknologi.v14i2.4650>
- Indriyani, I., Wiranata, I. P. B., & Hiu, S. (2024). Strategi Peningkatan Efisiensi Operasional UMKM di Era Digital: Pendekatan Kualitatif dengan Business Intelligence dalam Implementasi E-Commerce. *INFORMATICS FOR EDUCATORS AND PROFESSIONAL : Journal of Informatics*, 9(1), 23. <https://doi.org/10.51211/itbi.v9i1.2760>
- Nofitri, R., & Irawati, N. (2019). Analisis Data Hasil Keuntungan Menggunakan Software RapidMiner. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 5(2), 199–204. <https://doi.org/10.33330/jurteksi.v5i2.365>
- Pangaribuan, J. J., & Putra, A. (2024). Blood Donation Classification with Decision Tree Method using C4.5 Algorithm. *International Journal of Multidisciplinary Approach Research and Science*, 2(03), 1248–1259. <https://doi.org/10.59653/ijmars.v2i03.961>

- Putri, M. S. E., & Marsono, M. (2025). Analisis Peran Perantara terhadap Efisiensi Distribusi Tebu. *Lokawati : Jurnal Penelitian Manajemen Dan Inovasi Riset*, 3(4), 226–237. <https://doi.org/10.61132/lokawati.v3i4.1974>
- Resti, Y., Aryanto, R., Yahdin, S., & Kresnawati, E. S. (2023). Rain Event Prediction Performance Using Decision Tree Method. *AIP Conference Proceedings*, 2689(1), 120006. <https://doi.org/10.1063/5.0117434>
- Rusyana, N. R., Renaldi, F., & Destiani, D. (2023). Prediction Analysis of Four Disease Risk Using Decision Tree C4.5. *ICCoSITE 2023 - International Conference on Computer Science, Information Technology and Engineering: Digital Transformation Strategy in Facing the VUCA and TUNA Era*, 90–94. <https://doi.org/10.1109/ICCoSITE57641.2023.10127710>
- Saputra, Z. N., & Fatah, Z. (2025). Penggunaan Data Mining Untuk Mengidentifikasi Pelanggan Beresiko Tinggi Dalam Penjualan Menggunakan Algoritma Decision Tree C4.5. *JUSIFOR : Jurnal Sistem Informasi Dan Informatika*, 4(1), 68–74. <https://doi.org/10.70609/jusifor.v4i1.5942>
- Sibarani, H. A., Sarkis, I. M., & Simanullang, H. G. (2023). Penerapan Algoritma C4.5 Dalam Menentukan Pemberian Pinjaman Bagi Nasabah Di CU.Mitra Kasih Smart. *Methosisfo: Jurnal Ilmiah Sistem Informasi*, 3(1), 137–142. <http://ojs.fikom-methodist.net/index.php/methosisfo>
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi RapidMiner. *JBASE - Journal of Business and Audit Information Systems*, 4(1). <https://doi.org/10.30813/jbase.v4i1.2729>