

Model Klasifikasi *Decision Tree* dalam Menentukan Eligibilitas Nasabah Pinjaman Menggunakan Algoritma C4.5

¹Fachrizan Aditya, ²Zahira Putri Effia Yumna, ³Andre Rafli, ⁴Falah Rafif, ⁵Muhammad Ikhsan
^{1,2,3,4,5}Universitas Bina Sarana Informatika
Jakarta, Indonesia

¹fachrizan.ad@gmail.com, ²zahira.putri.ey@gmail.com, ³andrerafli83@gmail.com,
⁴falah.rfif@gmail.com, ⁵muhikhsan1108@gmail.com

*Penulis Korespondensi

Diajukan : 25/11/2025
Diterima : 18/12/2025
Dipublikasi : 01/01/2026

ABSTRAK

Penurunan suku bunga acuan sebesar 0,25% oleh *The Federal Reserve (The Fed)* memicu peningkatan permintaan kredit secara global. Kebijakan ini mendorong perputaran modal di sektor keuangan serta menimbulkan tantangan bagi pihak perbankan dalam menjaga perkembangan ekonomi global. Kondisi ini menuntut lembaga keuangan untuk dapat menilai kelayakan nasabah lebih bijak agar risiko kredit macet dapat diminimalisir. Penelitian ini bertujuan untuk membangun model klasifikasi menggunakan algoritma C4.5 dalam menentukan eligibilitas nasabah pinjaman kredit. Populasi dalam penelitian ini adalah calon nasabah yang mengajukan pinjaman di HDFC Bank Ltd, salah satu bank terbesar di India, dengan jumlah sampel acak sebanyak 3.192 entri. Metode yang digunakan adalah pendekatan kuantitatif dengan penerapan *data mining* berbasis *decision tree* melalui perangkat lunak RapidMiner. Hasil pengujian menunjukkan bahwa algoritma C4.5 berhasil membentuk model klasifikasi dengan variabel paling berpengaruh yaitu *Credit History* dengan nilai *information gain* sebesar 1,096930. Berdasarkan *confusion matrix*, model memperoleh akurasi 98%, dengan *recall* masing-masing 99,08% (*Approved*) dan 97,11% (*Rejected*). *Precision* yang dihasilkan adalah 96,85% untuk *Approved* dan 99,16% untuk *Rejected*. Berdasarkan hasil tersebut, algoritma C4.5 dapat membantu pihak perbankan dalam proses analisis kelayakan nasabah secara lebih objektif dan efisien, sehingga berpotensi mendukung perputaran ekonomi nasional melalui penyaluran kredit yang lebih tepat sasaran. Meskipun data yang digunakan merupakan simulasi dari lembaga keuangan di India, hasil penelitian ini diharapkan dapat menjadi referensi awal bagi penerapan sistem serupa pada lembaga keuangan serupa.

Kata Kunci: Algoritma C4.5, *Data Mining*, Klasifikasi, Pemberian Pinjaman, Penurunan Suku Bunga

I. PENDAHULUAN

Suku bunga merupakan instrumen utama yang digunakan bank sentral dalam mengelola stabilitas ekonomi. Dalam konteks ekonomi global, fluktuasi suku bunga menjadi perhatian utama, karena memiliki dampak yang luas terhadap berbagai sektor ekonomi (Ernani, 2024). Kenaikkan suku bunga akan meningkatkan biaya pinjaman yang berdampak pada turunnya minat untuk meminjam. Sebaliknya, penurunan suku bunga akan memudahkan akses pinjaman, yang dapat mendorong peningkatan konsumsi dan investasi, namun juga berpotensi meningkatkan tekanan inflasi karena permintaan yang lebih besar (Suseno & Agusalm, 2024).

Suku bunga kebijakan atau suku bunga acuan merupakan tingkat bunga yang ditetapkan oleh bank sentral untuk pinjaman jangka pendek dan menjadi alat utama dalam pelaksanaan kebijakan moneter di berbagai negara (Basmar et al., 2023). Salah satu contoh penerapan kebijakan tersebut adalah oleh *Federal Reserve System (The Fed)* sebagai bank sentral Amerika Serikat, yang

memiliki peran penting dalam mengarahkan stabilitas ekonomi global melalui penetapan *federal funds rate* (Wahid & Chaidir, 2025). Pada September 2024, *The Fed* mengeluarkan keputusan pemangkasan suku bunga acuan sebesar 0,25% untuk menanggapi perlambatan ekonomi dan melemahnya pasar tenaga kerja (Picchi, 2025). Kebijakan tersebut memicu peningkatan aktivitas kredit secara global dan menimbulkan tantangan bagi lembaga keuangan dalam menjaga kualitas portofolio kredit agar terhindar dari risiko kredit macet (Prasetyo et al., 2025). Untuk membantu lembaga keuangan dalam meminimalkan risiko kredit macet, penerapan *data mining* dapat menawarkan pendekatan inovatif untuk memanfaatkan data nasabah secara efektif dengan metode klasifikasi menggunakan algoritma *decision tree* C4.5 (Sari & Tegal, 2025).

Penelitian ini bertujuan untuk membangun model klasifikasi eligibilitas nasabah pinjaman menggunakan algoritma C4.5 yang merupakan salah satu metode *decision tree* dalam *data mining*. Algoritma ini dipilih kemampuannya menghasilkan aturan (*rule*) yang mudah dipahami dan akurat dalam mengklasifikasikan data. Populasi dari penelitian ini menggunakan data nasabah yang mengajukan pinjaman di HDFC Bank Ltd, salah satu bank terbesar di India, dengan jumlah sampel acak sebanyak 3.192 entri.

Diharapkan hasil penelitian ini dapat memberikan kontribusi bagi lembaga keuangan dalam meningkatkan akurasi pengambilan keputusan kredit agar meminimalisir risiko kredit macet serta menjadi referensi awal untuk pengembangan sistem penilaian kelayakan nasabah berbasis *data mining* di negara serupa.

II. STUDI LITERATUR

1. Penelitian Terdahulu

Penelitian mengenai penerapan algoritma C4.5 dalam menentukan eligibilitas nasabah pinjaman telah banyak dilakukan. Salah satunya adalah penelitian oleh (Pujiono et al., 2024) yang dilakukan pada Bank Sumsel Babel Cabang Pagar Alam dengan judul “Penerapan *Data Mining* untuk Mengklasifikasi Nasabah Bank Sumsel Babel Menggunakan Algoritma C4.5”. Penelitian ini bertujuan membantu *account officer* dalam mempercepat analisis kelayakan debitur KUR yang sebelumnya dilakukan secara manual dan memakan waktu. Metode algoritma C4.5 digunakan untuk mengklasifikasikan nasabah menjadi kategori lancar dan macet berdasarkan indikator seperti kondisi usaha, kecukupan jaminan, riwayat kredit, jumlah pinjaman, dan kemampuan membayar. Hasil penelitian menunjukkan akurasi model sebesar 80% dengan pohon keputusan yang dihasilkan mampu mendukung proses penilaian debitur secara lebih efektif dan efisien.

Penelitian lain dilakukan oleh (Wulan et al., 2023) dengan judul “Penerapan *Data Mining* dengan Algoritma C4.5 dalam Penentuan Kelayakan Nasabah Kredit” dilakukan untuk mengatasi proses penilaian kredit yang masih manual dan rawan kesalahan. Dengan menggunakan tahapan *CRISP-DM*, peneliti menerapkan algoritma C4.5 untuk mengklasifikasikan nasabah layak dan tidak layak berdasarkan data yang dikumpulkan. Hasil penelitian menunjukkan bahwa pendapatan menjadi atribut paling berpengaruh dalam pohon keputusan, di mana nasabah dengan pendapatan > Rp12.500.000 diprioritaskan dengan pengujian *confusion matrix* menghasilkan akurasi 81%.

Selain itu, penelitian yang dilakukan oleh (M. Zakaria & Andri, 2023) dengan judul “Penerapan *Data Mining* untuk Klasifikasi Kelayakan Nasabah dalam Pengajuan Kredit Menggunakan Metode *Naive Bayes* dan *Random Forest*” dilakukan untuk membantu bank menentukan kelayakan calon nasabah secara lebih akurat karena proses analisis pinjaman sebelumnya masih dilakukan secara manual. Peneliti menggunakan algoritma *Naive Bayes* dan *Random Forest* untuk mengklasifikasikan nasabah layak atau tidak layak berdasarkan parameter pinjaman yang tersedia. Hasil penelitian menunjukkan bahwa metode *Naive Bayes* memberikan akurasi sebesar 78,4%. Penerapan kedua algoritma ini dinilai mampu membantu bank dalam meminimalkan risiko kredit bermasalah serta meningkatkan ketepatan analisis kelayakan nasabah.

2. *Data Mining*

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis (Manullang et al., 2020). Dengan memanfaatkan *data mining*, lembaga keuangan dapat mengidentifikasi tren, pola nasabah, serta perilaku nasabah yang sebelumnya tidak terlihat

(Surojudin & Danny, 2025).

3. Algoritma C4.5

Algoritma C4.5 dikenal sebagai salah satu algoritma klasifikasi paling populer dalam bidang pembelajaran mesin karena kemampuannya menangani berbagai jenis data, baik numerik maupun kategorikal. Algoritma ini bekerja dengan membentuk pohon keputusan berdasarkan konsep entropi dan *gain* informasi untuk menentukan atribut terbaik dalam membagi data ke dalam kategori yang berbeda (Wirasena & Warmansyah, 2024). Secara umum, dalam membuat pohon keputusan menggunakan algoritma C4.5 memiliki beberapa tahap, yaitu (Kredit & Lembaga, 2024):

- A. Pemilihan atribut sebagai akar
- B. Buat cabang untuk tiap-tiap nilai
- C. Bagi kasus dalam cabang
- D. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

4. Pohon Keputusan (*Decision Tree*)

Algoritma yang akan mencari solusi permasalahan dengan menjadikan kriteria sebagai *node* yang saling berhubungan membentuk seperti struktur pohon adalah *decision tree*. Setiap pohon memiliki cabang, cabang mewakili suatu atribut yang harus dipenuhi untuk menuju cabang selanjutnya hingga berakhir di daun (Hidayatulloh et al., 2022). Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target (Manullang et al., 2020).

III. METODE

Penelitian ini dilakukan secara terstruktur menggunakan tahapan *Knowledge Discovery in Database* (KDD). Tahap ini menggunakan pendekatan sistematis untuk mengekstraksi pengetahuan dari data, meliputi pengumpulan data, *preprocessing* data, analisis data, dan interpretasi hasil.

1. Pengumpulan Data

Dataset yang digunakan merupakan data sintetik calon nasabah yang mengajukan pinjaman di HDFC Bank, yaitu salah satu bank paling terkenal di India. Dataset ini termasuk ke dalam data sekunder, yang diperoleh dari Kaggle (<https://www.kaggle.com/datasets/sanjamw/loan-approval-prediction-dataset>) dengan judul "*Loan Approval Prediction Dataset*". Dataset tersebut memiliki 3.192 *record* data dan 14 atribut awal dengan rincian sebagai berikut:

Tabel 1. Rincian Atribut Dataset

No	Atribut	Type Data	Keterangan
1.	Loan Id	Polynomial	ID Nasabah
2.	Gender	Binominal	Jenis Kelamin
3.	Married	Binominal	Status Perkawinan
4.	Dependents	Polynomial	Jumlah Tanggungan
5.	Education	Binominal	Latar Belakang Pendidikan
6.	Employment Status	Polynomial	Jenis Pekerjaan
7.	Applicant_Income	Real	Pendapatan Bulanan Nasabah (₹ Indian Rupee)
8.	Coapplicant_Income	Real	Pendapatan Bulanan Bersama, Jika Ada (₹ Indian Rupee)
9.	Loan_Amount	Real	Jumlah Pinjaman yang Diajukan (₹ Indian Rupee)
10.	Loan Term	Polynomial	Jangka Waktu Pinjaman (Bulan)
11.	Credit_History	Binominal	Indikator Riwayat Kredit (1 = Baik, 0 = Buruk)
12.	Property_Area	Polynomial	Jenis Lokasi Properti yang Dimiliki
13.	Age	Integer	Usia Nasabah
14.	Loan_Status	Binominal	Status Persetujuan Pinjaman

	(Label/Target)	
--	----------------	--

Tabel 2. Dataset Awal “Loan Approval Prediction”

No	Loan_ID	Gender	Married	Dependents	Education	Employment Status	Applicant Income	Coapplicant Income	Loan Amount	Loan Term	Credit History	Property Area	Age	Loan Status
1.	LN03437	Female	Yes	3	Graduate	Salaried	34.473,9	13.914,7	240.732,5	180	1	Semiurban	40.0	Approved
2.	LN04823	Female	Yes	1	Graduate	Self-Employed	59.718,0	18.720,8	122.547,3	180	1	Semiurban	47.2	Approved
3.	LN04186	Female	Yes	1	Not Graduate	Salaried	106.981,5	2.229,2	140.550,9	60	1	Rural	24.2	Approved
4.	LN00380	Male	Yes	2	Graduate	Salaried	43.205,6	35.578,1	146.135,7	120	1	Semiurban	38.7	Approved
5.	LN04698	Female	Yes	0	Graduate	Salaried	62.526,3	20.212,2	203.408,1	180	1	Semiurban	41.1	Approved
3188.	LN02182	Female	Yes	0	Graduate	Self-Employed	56.392,2	14.981,7	140.736,7	360	1	Urban	32.4	Approved
3189.	LN05878	Male	Yes	2	Graduate	Salaried	43.651,8	17.992,0	67.482,3	180	1	Rural	41.4	Approved
3190.	LN03290	Male	Yes	0	Not Graduate	Salaried	88.370,6	8.276,1	113.500,2	300	1	Rural	45.2	Approved
3191.	LN00770	Male	No	1	Not Graduate	Salaried	74.155,0	14.679,5	214.897,0	360	1	Urban	26.6	Approved
3192.	LN05005		Yes	1	Graduate	Salaried	34.515,4	12.881,1	69.015,5	360	0	Semiurban	38.8	Rejected

2. Data Preparation

Proses data preparation pada dataset ini meliputi:

- 1) Pemilihan Prediktor
- 2) Penghapusan Data Missing Value
- 3) Pembagian Data Training & Data Testing

3. Implementasi Algoritma C4.5

Penggunaan algoritma C4.5 pada pemodelan dataset ini bertujuan menghasilkan pohon keputusan untuk membantu menilai kelayakan pinjaman kepada nasabah secara lebih objektif, akurat, dan efisien (Hidayatulloh et al., 2022).

- 1) Perhitungan Entropy dan Information Gain

Untuk menghitung nilai entropy dapat menggunakan rumus berikut:

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

Keterangan:

- S: Jumlah seluruh data
- n: Jumlah kumpulan data
- pi: Nilai atribut dibagi label hasil
- log2: Logaritma biner / basis dua

Untuk menghitung information gain dapat menggunakan rumus berikut:

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

Keterangan:

- S: Jumlah seluruh data
- A: Atribut atau kolom
- n: Jumlah partisi S
- |Si|: Jumlah kasus pada partisi ke-i
- |S|: Jumlah kasus dalam S

- 2) Decision Tree



Pohon keputusan (*decision tree*) merupakan hasil dari proses perhitungan *entropy* dan *information gain* yang telah dilakukan di mana nilai *information gain* yang tertinggi akan dijadikan sebagai akar (*root*) dari *decision tree* tersebut.

4. Rule Model

Rule model merupakan penjelasan detail yang mendeskripsikan sebuah pohon keputusan dalam bentuk logika yang mudah dibaca.

5. Evaluasi dan Validasi Model

Berikut beberapa ukuran evaluasi utama yang akan diperhatikan:

Tabel 3. Ukuran Evaluasi Utama

Metriks	Fungsi	Rumus	Arti
<i>Confusion Matrix</i>	Menunjukkan distribusi hasil prediksi	-	Menampilkan TP, TN, FP, FN secara tabel
<i>Accuracy</i>	Mengukur ketepatan prediksi model secara keseluruhan	$(TP + TN) / (TP + TN + FP + FN)$	Persentase prediksi benar dari semua data
<i>Precision</i>	Mengukur ketepatan model saat memprediksi positif	$TP / (TP + FP)$	Dari semua yang diprediksi "Approved", berapa yang benar-benar "Approved"
<i>Recall</i>	Mengukur seberapa banyak kasus positif yang berhasil ditemukan	$TP / (TP + FN)$	Dari semua "Approved" sebenarnya, berapa yang terdeteksi model

IV. HASIL DAN PEMBAHASAN

1. Data Preparation

1) Pemilihan Prediktor

Pada dataset ini perlu adanya penghapusan atribut-atribut yang berpotensi menyebabkan bias atau diskriminasi dalam model, diantaranya adalah atribut *Gender*, *Age*, *Marriage*, dan *Education*, lalu akan dihilangkan juga atribut *Loan_Id* karena tidak ada relevansinya pada penerapan algoritma C4.5. Adapun atribut-atribut yang digunakan sebagai prediktor adalah:

Tabel 4. Atribut yang Digunakan

No	Atribut Terpilih	Keterangan
1.	Dependents	Prediktor
2.	Property Area	Prediktor
3.	Employment Status	Prediktor
4.	Credit History	Prediktor
5.	Loan Term	Prediktor
6.	Applicant Income	Prediktor
7.	Coapplicant Income	Prediktor
8.	Loan Amount	Prediktor
9.	Loan Status	Label

2) Penghapusan Data *Missing Value*

Pada dataset ini terdapat data yang *missing value* sebanyak 195 data. Dengan melakukan penghapusan data *missing value* pada dataset ini, maka didapat jumlah data yang *no missing value* adalah sebanyak 3063 data. Berikut merupakan data akhir yang akan digunakan dalam algoritma C4.5 dengan menggunakan 9 atribut dan 3063 data:

Tabel 5. Dataset Calon Nasabah Kredit

No	Dependents	Property Area	Employment Status	Credit History	Loan Term	Applicant Income	Coapplicant Income	Loan Amount	Loan Status
1.	3	Semiurban	Salaried	1	180	34.473,9	13.914,7	240.732,5	Approved
2.	1	Semiurban	Self-Employed	1	180	59.718,0	18.720,8	122.547,3	Approved
3.	1	Rural	Salaried	1	60	106.981,5	2.229,2	140.550,9	Approved
4.	2	Semiurban	Salaried	1	120	43.205,6	35.578,1	146.135,7	Approved
5.	0	Semiurban	Salaried	1	180	62.526,3	20.212,2	203.408,1	Approved
3059.	0	Urban	Self-Employed	1	360	56.392,2	14.981,7	140.736,7	Approved
3060.	2	Rural	Salaried	1	180	43.651,8	17.992,0	67.482,3	Approved
3061.	0	Rural	Salaried	1	300	88.370,6	8.276,1	113.500,2	Approved
3062.	1	Urban	Salaried	1	360	74.155,0	14.679,5	214.897,0	Approved
3063.	1	Semiurban	Salaried	0	360	34.515,4	12.881,1	69.015,5	Rejected

3) Pembagian *Data Training & Data Testing*

Pada dataset ini akan dilakukan pembagian data secara acak menjadi *data training* untuk melatih model dan *data testing* untuk menguji performa model dengan komposisi 70% untuk *data training* dan 30% untuk *data testing*. Setelah dilakukan *split data*, maka dataset akan terbagi ke dalam *data training* sebanyak 2.144 data dan *data testing* sebanyak 919 data.

2. Implementasi Algoritma C4.5

1) Perhitungan *Entropy* dan *Information Gain*

Berikut adalah tabel hasil perhitungan *entropy* dan *gain* yang dilakukan pada *data training*:

Tabel 6. Hasil Perhitungan *Entropy* dan *Information Gain*

Node	Keterangan	Total	Approved	Rejected	Entropy	Gain
Total		2144	1100	1044	0,999508	
Dependents	0	808	409	399	0,999890	0,000485
	1	682	343	339	0,999975	
	2	454	243	211	0,996413	
	3	200	105	95	0,998196	
Property Area	Rural	421	201	220	0,998530	0,000922
	Semiurban	634	328	306	0,999131	
	Urban	1089	571	518	0,998291	
Employment Status	Salaried	1494	797	697	0,996766	0,002823
	Self-Employed	549	254	295	0,995973	
	Unemployed	101	49	52	0,999363	
Credit History	0	1011	1	1010	0,011299	1,096930
	1	1133	1099	34	0,194436	
Loan Term	60	319	172	147	0,995565	0,001357
	120	352	169	183	0,998859	
	180	364	198	166	0,994418	
	240	353	178	175	0,999948	
	300	388	197	191	0,999827	
	360	368	186	182	0,999915	
Applicant Income	< 20.000	50	18	32	0,942683	0,001738
	> 20.000	613	312	301	0,999768	
	> 50.000	1435	745	690	0,998940	
	> 100.000	46	25	21	0,994539	
Coapplicant Income	0	149	76	73	0,999708	0,000485
	< 10.000	576	293	283	0,999783	
	> 10.000	1268	660	608	0,998787	
	> 30.000	151	71	80	0,997436	
Loan Amount	< 50.000	50	26	24	0,998846	0,003124
	> 50.000	283	134	149	0,997973	

	> 100.000	1447	774	673	0,996483	
	> 200.000	364	166	198	0,994418	

Berdasarkan tabel perhitungan di atas, berikut adalah uraian cara menghitung *entropy* dan *information gain*:

a. Menghitung *Entropy*

1. *Entropy Total*

$$= ((-1100/2144) * \log_2(1100/2144)) + ((-1044/2144) * \log_2(1044/2144)) = 0,999508$$
2. *Entropy Dependents (0)*

$$= ((-409/808) * \log_2(409/808)) + ((-399/808) * \log_2(399/808)) = 0,999890$$
3. *Entropy Dependents (1)*

$$= ((-343/682) * \log_2(343/682)) + ((-339/682) * \log_2(339/682)) = 0,999975$$
4. *Entropy Dependents (2)*

$$= ((-243/454) * \log_2(243/454)) + ((-211/454) * \log_2(211/454)) = 0,996413$$
5. *Entropy Dependents (3)*

$$= ((-105/200) * \log_2(105/200)) + ((-95/200) * \log_2(95/200)) = 0,998196$$
6. *Entropy Property Area (Rural)*

$$= ((-201/421) * \log_2(201/421)) + ((-220/421) * \log_2(220/421)) = 0,998530$$
7. *Entropy Property Area (Semiurban)*

$$= ((-328/634) * \log_2(328/634)) + ((-306/634) * \log_2(306/634)) = 0,999131$$
8. *Entropy Property Area (Urban)*

$$= ((-571/1089) * \log_2(571/1089)) + ((-518/1089) * \log_2(518/1089)) = 0,998291$$
9. *Entropy Employment Status (Salaried)*

$$= ((-797/1494) * \log_2(797/1494)) + ((-697/1494) * \log_2(697/1494)) = 0,996766$$
10. *Entropy Employment Status (Self-Employed)*

$$= ((-254/549) * \log_2(254/549)) + ((-295/549) * \log_2(295/549)) = 0,995973$$
11. *Entropy Employment Status (Unemployed)*

$$= ((-49/101) * \log_2(49/101)) + ((-52/101) * \log_2(52/101)) = 0,999363$$
12. *Entropy Credit History (0)*

$$= ((-1/1011) * \log_2(1/1011)) + ((-1010/1011) * \log_2(1010/1011)) = 0,011299$$
13. *Entropy Credit History (1)*

$$= ((-1099/1133) * \log_2(1099/1133)) + ((-34/1133) * \log_2(34/1133)) = 0,194436$$
14. *Entropy Loan Term (60)*

$$= ((-172/319) * \log_2(172/319)) + ((-147/319) * \log_2(147/319)) = 0,995565$$
15. *Entropy Loan Term (120)*

$$= ((-169/352) * \log_2(169/352)) + ((-183/352) * \log_2(183/352)) = 0,998859$$
16. *Entropy Loan Term (180)*

$$= ((-198/364) * \log_2(198/364)) + ((-166/364) * \log_2(166/364)) = 0,994418$$
17. *Entropy Loan Term (240)*

$$= ((-178/353) * \log_2(178/353)) + ((-175/353) * \log_2(175/353)) = 0,999948$$
18. *Entropy Loan Term (300)*

$$= ((-197/388) * \log_2(197/388)) + ((-191/388) * \log_2(191/388)) = 0,999827$$
19. *Entropy Loan Term (360)*

$$= ((-186/368) * \log_2(186/368)) + ((-182/368) * \log_2(182/368)) = 0,999915$$
20. *Entropy Applicant Income (< 20.000)*

$$= ((-18/50) * \log_2(18/50)) + ((-32/50) * \log_2(32/50)) = 0,942683$$
21. *Entropy Applicant Income (> 20.000)*

$$= ((-312/613) * \log_2(312/613)) + ((-301/613) * \log_2(301/613)) = 0,999768$$
22. *Entropy Applicant Income (> 50.000)*

$$= ((-745/1435) * \log_2(745/1435)) + ((-690/1435) * \log_2(690/1435)) = 0,99894$$
23. *Entropy Applicant Income (> 100.000)*

$$= ((-25/46) * \log_2(25/46)) + ((-21/46) * \log_2(21/46)) = 0,994539$$
24. *Entropy Coapplicant Income (0)*

$$= ((-76/149) * \log_2 (76/149)) + ((-73/149) * \log_2 (73/149)) = 0,999708$$

25. *Entropy Coapplicant Income (< 10.000)*

$$= ((-293/576) * \log_2 (293/576)) + ((-283/576) * \log_2 (283/576)) = 0,999783$$

26. *Entropy Coapplicant Income (> 10.000)*

$$= ((-660/1268) * \log_2 (660/1268)) + ((-608/1268) * \log_2 (608/1268)) = 0,998787$$

27. *Entropy Coapplicant Income (> 30.000)*

$$= ((-71/151) * \log_2 (71/151)) + ((-80/151) * \log_2 (80/151)) = 0,997436$$

28. *Entropy Loan Amount (< 50.000)*

$$= ((-26/50) * \log_2 (26/50)) + ((-24/50) * \log_2 (24/50)) = 0,998846$$

29. *Entropy Loan Amount (> 50.000)*

$$= ((-134/283) * \log_2 (134/283)) + ((-149/283) * \log_2 (149/283)) = 0,997973$$

30. *Entropy Loan Amount (> 100.000)*

$$= ((-774/1447) * \log_2 (774/1447)) + ((-673/1447) * \log_2 (673/1447)) = 0,996483$$

31. *Entropy Loan Amount (> 200.000)*

$$= ((-166/364) * \log_2 (166/364)) + ((-198/364) * \log_2 (198/364)) = 0,994418$$

b. Menghitung *Information Gain*

1. *Gain Dependents*

$$= (0,999508) - ((808/2144) * 0,99989) - ((682/2144) * 0,999975) - ((454/2144) * 0,996413) - ((200/2144) * 0,998196) = 0,000485$$

2. *Gain Property Area*

$$= (0,999508) - ((421/2144) * 0,99853) - ((634/2144) * 0,999131) - ((1089/2144) * 0,998291) = 0,000922$$

3. *Gain Employment Status*

$$= (0,999508) - ((1494/2144) * 0,996766) - ((549/2144) * 0,995973) - ((101/2144) * 0,999363) = 0,002823$$

4. *Gain Credit History*

$$= (0,999508) - ((1011/2144) * 0,011299) - ((1133/2144) * 0,194436) = 1,096930$$

5. *Gain Loan Term*

$$= (0,999508) - ((319/2144) * 0,995565) - ((352/2144) * 0,998859) - ((364/2144) * 0,994418) - ((353/2144) * 0,999948) - ((388/2144) * 0,999827) - ((368/2144) * 0,999915) = 0,001357$$

6. *Gain Applicant Income*

$$= (0,999508) - ((50/2144) * 0,942683) - ((613/2144) * 0,999768) - ((1435/2144) * 0,99894) - ((46/2144) * 0,994539) = 0,001738$$

7. *Gain Coapplicant Income*

$$= (0,999508) - ((149/2144) * 0,999708) - ((576/2144) * 0,999783) - ((1268/2144) * 0,998787) - ((151/2144) * 0,997436) = 0,000485$$

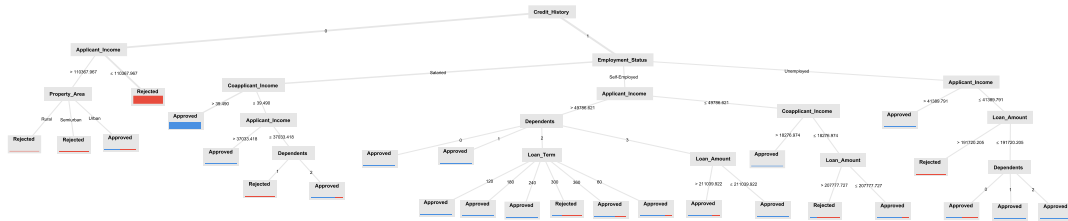
8. *Gain Loan Amount*

$$= (0,999508) - ((50/2144) * 0,998846) - ((283/2144) * 0,997973) - ((1447/2144) * 0,996483) - ((364/2144) * 0,994418) = 0,003124$$

Berdasarkan perhitungan *information gain* yang dilakukan, diperoleh atribut dengan nilai *gain* tertinggi adalah *Credit History* yaitu sebesar 1,096930. Maka atribut *Credit History* dapat ditentukan menjadi *node* (akar) dari model *decision tree*. Atribut *Credit History* memiliki 2 klasifikasi, yakni 0 dan 1, yang menandakan bahwa nilai 0 berarti memiliki indikator riwayat kredit yang buruk, sedangkan nilai 1 memiliki indikator riwayat kredit yang baik.

2) *Decision Tree*

Berikut adalah gambar pohon keputusan (*decision tree*) yang dibuat setelah proses perhitungan *entropy* dan *information gain* pada *data training*:



Gambar 1. Pohon Keputusan Eligibilitas Pinjaman Kredit

Berdasarkan gambar pohon keputusan di atas, *Credit History* berperan sebagai *node* yang membagi menjadi 2 klasifikasi yaitu 0 & 1, dan terdapat beberapa *leaf* atau kondisi-kondisi yang menghasilkan nilai *Approved* dan juga *Rejected*. *Credit History* menjadi faktor paling kuat dalam menentukan kelayakan pinjaman.

3. Rule Model

Berdasarkan pohon keputusan yang terbentuk, maka terdapat aturan atau *rule model* yang juga terbentuk dengan rincian sebagai berikut:

1) Untuk *Credit_History* = 0

1. IF *Credit_History* = 0 AND *Applicant_Income* ≤ 110367.967
 THEN *Loan_Status* = *Rejected*
2. IF *Credit_History* = 0 AND *Applicant_Income* > 110367.967 AND *Property_Area* = *Rural*
 THEN *Loan_Status* = *Rejected*
3. IF *Credit_History* = 0 AND *Applicant_Income* > 110367.967 AND *Property_Area* = *Semiurban*
 THEN *Loan_Status* = *Rejected*
4. IF *Credit_History* = 0 AND *Applicant_Income* > 110367.967 AND *Property_Area* = *Urban*
 THEN *Loan_Status* = *Approved*

2) Untuk *Credit_History* = 1

5. IF *Credit_History* = 1 AND *Employment_Status* = *Salaried* AND *Coapplicant_Income* > 39.490
 THEN *Loan_Status* = *Approved*
6. IF *Credit_History* = 1 AND *Employment_Status* = *Salaried* AND *Coapplicant_Income* ≤ 39.490 AND *Applicant_Income* > 37033.418
 THEN *Loan_Status* = *Approved*
7. IF *Credit_History* = 1 AND *Employment_Status* = *Salaried* AND *Coapplicant_Income* ≤ 39.490 AND *Applicant_Income* ≤ 37033.418 AND *Dependents* = 1
 THEN *Loan_Status* = *Rejected*
8. IF *Credit_History* = 1 AND *Employment_Status* = *Salaried* AND *Coapplicant_Income* ≤ 39.490 AND *Applicant_Income* ≤ 37033.418 AND *Dependents* = 2
 THEN *Loan_Status* = *Approved*
9. IF *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* > 49786.621 AND *Dependents* = 0
 THEN *Loan_Status* = *Approved*
10. IF *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* > 49786.621 AND *Dependents* = 1
 THEN *Loan_Status* = *Approved*
11. IF *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND

- Applicant_Income* > 49786.621 AND *Dependents* = 2 AND *Loan_Term* = 60
THEN *Loan_Status* = *Approved*
12. **IF** *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* > 49786.621 AND *Dependents* = 2 AND *Loan_Term* = 120
THEN *Loan_Status* = *Approved*
13. **IF** *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* > 49786.621 AND *Dependents* = 2 AND *Loan_Term* = 180
THEN *Loan_Status* = *Approved*
14. **IF** *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* > 49786.621 AND *Dependents* = 2 AND *Loan_Term* = 240
THEN *Loan_Status* = *Approved*
15. **IF** *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* > 49786.621 AND *Dependents* = 2 AND *Loan_Term* = 300
THEN *Loan_Status* = *Rejected*
16. **IF** *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* > 49786.621 AND *Dependents* = 2 AND *Loan_Term* = 360
THEN *Loan_Status* = *Approved*
17. **IF** *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* > 49786.621 AND *Dependents* = 3 AND *Loan_Amount* > 211039.922
THEN *Loan_Status* = *Approved*
18. **IF** *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* > 49786.621 AND *Dependents* = 3 AND *Loan_Amount* ≤ 211039.922
THEN *Loan_Status* = *Approved*
19. **IF** *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* ≤ 49786.621 AND *Coapplicant_Income* > 18276.974
THEN *Loan_Status* = *Approved*
20. **IF** *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* ≤ 49786.621 AND *Coapplicant_Income* ≤ 18276.974 AND *Loan_Amount* > 207777.727
THEN *Loan_Status* = *Rejected*
21. **IF** *Credit_History* = 1 AND *Employment_Status* = *Self-Employed* AND *Applicant_Income* ≤ 49786.621 AND *Coapplicant_Income* ≤ 18276.974 AND *Loan_Amount* ≤ 207777.727
THEN *Loan_Status* = *Approved*
22. **IF** *Credit_History* = 1 AND *Employment_Status* = *Unemployed* AND *Applicant_Income* > 41389.791
THEN *Loan_Status* = *Approved*
23. **IF** *Credit_History* = 1 AND *Employment_Status* = *Unemployed* AND *Applicant_Income* ≤ 41389.791 AND *Loan_Amount* > 191720.205
THEN *Loan_Status* = *Rejected*
24. **IF** *Credit_History* = 1 AND *Employment_Status* = *Unemployed* AND *Applicant_Income* ≤ 41389.791 AND *Loan_Amount* ≤ 191720.205 AND *Dependents* = 0
THEN *Loan_Status* = *Approved*
25. **IF** *Credit_History* = 1 AND *Employment_Status* = *Unemployed* AND *Applicant_Income* ≤ 41389.791 AND *Loan_Amount* ≤ 191720.205 AND *Dependents* = 1
THEN *Loan_Status* = *Approved*
26. **IF** *Credit_History* = 1 AND *Employment_Status* = *Unemployed* AND *Applicant_Income* ≤ 41389.791 AND *Loan_Amount* ≤ 191720.205 AND *Dependents* = 2
THEN *Loan_Status* = *Approved*

Berdasarkan *rule* di atas, dapat disimpulkan bahwa atribut *Credit_History* menjadi faktor paling kuat dalam menentukan kelayakan pinjaman. Nasabah dengan *Credit_History* = 1 (baik) hampir selalu *Approved*, kecuali pada kombinasi kondisi tertentu (*Applicant_Income* rendah & *Loan_Amount* besar). Sedangkan nasabah dengan *Credit_History* = 0 (buruk) hampir selalu

Rejected, meskipun *Applicant_Income* nya tinggi.

4. Evaluasi dan Validasi Model

Evaluasi dan validasi model klasifikasi dilakukan menggunakan *data testing* sebanyak 919 data. Evaluasi dan validasi model ini akan menggunakan *confusion matrix* sebagai berikut:

accuracy: 98.04%			
	true Approved	true Rejected	class precision
pred. Approved	431	14	96.85%
pred. Rejected	4	470	99.16%
class recall	99.08%	97.11%	

Gambar 2. Hasil Evaluasi dan Validasi Model menggunakan *Confusion Matrix*

Berdasarkan *confusion matrix* di atas, pengujian ini menghasilkan tingkat *accuracy* dalam memprediksi eligibilitas nasabah pinjaman kredit sebesar 98,04%, *recall* pada nilai *Approved* sebesar 99,08% dan nilai *Rejected* sebesar 97,11%, serta *precision* untuk prediksi *Approved* sebesar 96,85% dan untuk prediksi *Rejected* sebesar 99,16%. Model berhasil memprediksi dengan benar nilai *Approved* sebagai *Approved* sebanyak 431 kali (**TP**), 14 kali salah dalam memprediksi *Approved* sebagai *Rejected* (**FP**), 4 kali salah memprediksi *Rejected* sebagai *Approved* (**FN**), dan berhasil memprediksi dengan benar nilai *Rejected* sebagai *Rejected* (**TN**) sebanyak 470 kali.

V. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan, metode klasifikasi pohon keputusan (*decision tree*) dengan algoritma C4.5 yang diproses menggunakan RapidMiner dapat digunakan dengan baik dan efektif dalam menentukan eligibilitas nasabah pinjaman kredit di HDFC Bank Ltd. Model yang dihasilkan mampu menghasilkan visualisasi struktur pohon keputusan dengan jelas sehingga mudah dipahami oleh pihak analisis maupun pengambil keputusan. Hasil pengujian menunjukkan bahwa algoritma C4.5 berhasil membentuk model klasifikasi dengan variabel paling berpengaruh yaitu *Credit_History* dengan nilai *information gain* sebesar 1,096930. Berdasarkan *confusion matrix*, model memperoleh akurasi 98%, dengan *recall* masing-masing 99,08% (*Approved*) dan 97,11% (*Rejected*). *Precision* yang dihasilkan adalah 96,85% untuk *Approved* dan 99,16% untuk *Rejected*. Dengan demikian, dapat disimpulkan bahwa model klasifikasi algoritma C4.5 sangat direkomendasikan untuk digunakan oleh perusahaan perbankan dalam mengklasifikasikan seorang nasabah disetujui atau tidak dalam permintaan pinjaman kreditnya.

VI. REFERENSI

- Basmar, E., Hasan, M., & Astuty, S. (2023). *EKONOMI MONOTER* (T. Media (ed.)). CV TAHTA MEDIA GROUP.
- Ernani, A. I. (2024). *Pengaruh Suku Bunga terhadap Perkembangan Ekonomi Nasional Aulia*. 502–509.
- Hidayatulloh, T., Fajria, A., Lestari, R. N., & Nufus, N. S. Z. (2022). Algoritma C4.5 Untuk Menentukan Kelayakan Pemberian Kredit (Studi kasus: Bank Mandiri Taspen Kantor Kas Sukabumi). *Jurnal Larik: Ladang Artikel Ilmu Komputer*, 2(2), 66–74. <https://doi.org/10.31294/larik.v2i2.1836>
- Kredit, R., & Lembaga, P. (2024). *Tsalatsatun Nur Rohmah, Kurniawati PENERAPAN DATA MINING ALGORITMA C4.5 DALAM MENGIDENTIFIKASI RISIKO KREDIT PADA LEMBAGA KEUANGAN 1,2*. 2(1), 9–16.
- M. Zakaria, & Andri. (2023). *Penerapan Data Mining Untuk Klasifikasi Kelayakan Nasabah Dalam Pengajuan Kredit Menggunakan Metode Naives Bayes dan Random Forest*. 06(03), 391–397.

- Manullang, S. D., Buulolo, E., & Lubis, I. (2020). Implementasi Data Mining Dalam Memprediksi Jumlah Pinjaman Dengan Algoritma C4.5 Pada Kopdit CU Damai Sejahtera. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 1(3), 265. <https://doi.org/10.30865/json.v1i3.2153>
- Picchi, A. (2025). *Federal Reserve cuts interest rates by 0.25 percentage points amid weaker labor market*. Cbsnews.Com. <https://www.cbsnews.com/news/federal-reserve-interest-rate-decision-today-fomc-meeting/>
- Prasetyo, I., Sundoro, F. M., Amalia, F., Aini, N., & Anjarwani, F. (2025). *Menelusuri Evolusi Penelitian Quantitative Easing : Analisis Bibliometrik atas Respons terhadap Krisis Global*. 23(3), 458–474. <https://doi.org/10.26623/slsi.v23i3.12436>
- Pujiono, S., Astuti, R., & Basysyar, F. M. (2024). *IMPLEMENTASI DATA MINING UNTUK MENENTUKAN POLA PENJUALAN PRODUK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING*. 8(1).
- Sari, M. M., & Tegal, U. P. (2025). *Klasifikasi Data Nasabah Kredit Menggunakan Data Mining Dengan Algoritma Decission Tree*. 3(3), 5095–5100.
- Surojudin, N., & Danny, M. (2025). *Jurnal Informatika Ekonomi Bisnis Penerapan Data Mining dengan Algoritma C4 . 5 dan K-nearest Neighbor untuk Prediksi Penjualan Bahan Bangunan Terlaris*. 7, 672–679. <https://doi.org/10.37034/infeb.v7i3.1241>
- Suseno, A. M., & Aguslim, L. (2024). Hubungan Suku Bunga dan Inflasi: Studi Empiris di Indonesia. *Journal of Economic, Management and Entrepreneurship*, 2(3), 122–131. <https://doi.org/10.61502/jemes.v2i3.104>
- Wahid, M. R. J., & Chaidir, T. (2025). *Pengaruh Suku Bunga Federal Reserve System Dan Nilai Tukar Terhadap Inflasi Di Indonesia*. 7(1). <https://doi.org/10.32877/ef.v7i1.2157>
- Wirasena, M. R., & Warmansyah, J. (2024). Penerapan Algoritma C4.5 Untuk Prediksi Kelayakan Pengajuan Kartu Kredit Visa Bagi Nasabah. *TeknoIS : Jurnal Ilmiah Teknologi Informasi Dan Sains*, 14(2), 296–302. <https://doi.org/10.36350/jbs.v14i2.266>
- Wulan, A. A. A., Mahendra, I. P., Wardana, A., & Oky, K. (2023). PENERAPAN DATA MINING DENGAN ALGORITMA C4.5 DALAM PENENTUAN KELAYAKAN NASABAH KREDIT (STUDI KASUS: LPD DESA ADAT SUMERTA). *RESI Jurnal Riset Sistem Informasi*, 120–128. <https://doi.org/10.32795/resi.v1i3.4265>