

Perbandingan Metode Klasifikasi Naïve Bayes dan C4.5 untuk Menentukan Potensi Nasabah Pada NSC Finance

Marhaeni¹, Eviana Tjatur Putri², Roman Gusmana³

^{1,2,3}Program Studi Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati, Tarakan, Kalimantan Utara
Email: ¹marhaeni.reni1299@gmail.com, ²eviana@ppkia.ac.id, ³roman@ppkia.ac.id

Abstrak

NSC Finance, usaha jasa yang memberikan pinjaman kepada masyarakat untuk memenuhi kebutuhan. Masalah utama dari NSC Finance yaitu belum memanfaatkan data nasabahnya untuk mendapatkan informasi yang penting. Tujuan dari penelitian ini melakukan klasifikasi pada data nasabah untuk memperoleh informasi target nasabah mana yang memiliki berpotensi, dipertimbangkan, dan tidak berpotensi untuk penawaran pinjaman kembali. Untuk membantu sistem klasifikasi nasabah lebih akurat, ada beberapa metode yang dapat dibandingkan di antaranya yaitu Naïve Bayes dan C4.5 dengan pengukuran akurasi menggunakan recall precision. Analisa Perbandingan metode ini bertujuan untuk mengetahui metode yang memiliki akurasi tertinggi dalam menentukan klasifikasinya. Dengan adanya perbandingan metode ini perusahaan bisa mengetahui keakuratan tertinggi dalam menentukan hasil klasifikasi diantara 2 metode yang dibandingkan. Hasil penelitian yang ditemukan bahwa proses klasifikasi berdasarkan pola yang terbentuk dari 80 data training dan 20 data testing sehingga memungkinkan masih ada data yang klasifikasinya tidak bernilai sama dengan data aslinya, dan perbandingan metode klasifikasi data mining Naïve Bayes dan C4.5 menunjukkan bahwa Naïve Bayes lebih akurat dari pada C4.5. Hal tersebut dibuktikan dengan nilai akurasi dari Naïve Bayes yaitu 85%, precision 94.44%, dan recall 89.47%.

Kata Kunci: Perbandingan, Naïve Bayes, C4.5, Klasifikasi.

Comparison of Classification Naïve Bayes and C4.5 for NSC Finance's Potential Customers Decision

Abstract

NSC Finance is a service business that provides loans to the public to meet their needs. However, NSC Finance does not use customer data to obtain necessary information. This research classifies customer data to gain information about promising customers, considered customers, and unpromising customers to loan re-offer. This study compares Naïve Bayes and C4.5 to help customer classification systems be more accurate by measuring accuracy using recall precision. These methods' comparative analyses are to investigate which methods have the highest classification accuracy. Therefore, the company can discover the highest accuracy rate of the classification results of these two methods. Results revealed that the classification patterns of 80 training data and 20 test data make it possible that data still have classification differences from the original data. Methods comparison indicated that the Naïve Bayes classification is better, with 85% accuracy, 94.44% precision, and 89.47% recall.

Keywords: Comparison, Naïve Bayes, C4.5, Classification.

I. PENDAHULUAN

Evolusi data tidak terlepas dari perkembangan teknologi informasi saat ini, yang mampu mengumpulkan sejumlah besar data dalam kurun waktu tertentu. Data yang telah terkumpul dapat memberikan informasi berharga melalui proses penambangan data, yang dikenal sebagai *Data Mining*. Salah satu teknik dalam data mining yang banyak dimanfaatkan adalah klasifikasi data. Klasifikasi data adalah proses memisahkan data yang heterogen ke dalam kelompok

data yang homogen, sehingga sifat-sifat menonjol dari data tersebut dapat dengan mudah diidentifikasi.

Salah satu sektor jasa yang memiliki banyak data karena aktivitasnya yang dibutuhkan masyarakat saat ini adalah sektor keuangan. Finance, yang mencakup pengelolaan dan penyediaan dana untuk memenuhi kebutuhan calon nasabah atau konsumen, meliputi aktivitas seperti perbankan, utang, kredit, pasar modal, uang, dan investasi.

NSC Finance, sebagai sebuah usaha jasa, saat ini belum memanfaatkan data nasabahnya untuk mendapatkan informasi yang penting bagi pengelolanya. Penelitian ini bertujuan untuk melakukan klasifikasi terhadap data nasabah guna memperoleh informasi mengenai target nasabah yang berpotensi, perlu dipertimbangkan, atau tidak berpotensi untuk penawaran pinjaman kembali, dengan menganalisis data nasabah yang ada. Algoritma klasifikasi yang digunakan dalam data mining antara lain adalah Naïve Bayes dan C4.5.

Naïve Bayes adalah metode klasifikasi yang menggunakan prinsip probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris, Thomas Bayes. Algoritma Naïve Bayes memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu, dikenal sebagai Teorema Bayes. Ciri utama dari Naïve Bayes Classifier adalah asumsi independensi yang sangat kuat antara setiap kondisi atau kejadian.

Penelitian yang dilakukan oleh Karlana Indriani dan Qonita Tanjung pada tahun 2018 menunjukkan bahwa dengan menggunakan sistem pendukung keputusan berbasis metode Naïve Bayes untuk menentukan kelayakan pengajuan kredit motor, perusahaan dapat memproses data pendaftar dengan cepat dan efisien [1]. Sistem tersebut mengolah 15.625 data training dan 100 data testing yang dipilih secara acak, dan hasil pengujian menunjukkan tingkat akurasi sebesar 99% serta error sebesar 1%, sesuai dengan hasil yang diperoleh melalui aplikasi Rapidminer.

C4.5 adalah algoritma dalam kelompok *decision tree*. Algoritma ini menggunakan input berupa training samples dan samples. Training samples adalah data contoh yang digunakan untuk membangun tree dan telah diuji kebenarannya, sedangkan samples adalah field-field data yang digunakan sebagai parameter dalam proses klasifikasi data.

Penelitian lainnya oleh Larissa Navia Rani pada tahun 2016 menunjukkan bahwa algoritma C4.5 sangat efektif dalam klasifikasi data karena dapat menyajikan karakteristik data dengan jelas [2]. Algoritma ini memungkinkan data diklasifikasikan baik melalui struktur pohon keputusan (*decision tree*) maupun aturan If-Then, yang memudahkan pengguna dalam menggali informasi dari data yang bersangkutan. Temuan ini mendukung pemahaman bahwa C4.5 tidak hanya meningkatkan akurasi klasifikasi, tetapi juga menghasilkan hasil yang mudah diinterpretasikan, sehingga sangat bermanfaat dalam proses penggalian informasi.

Pada tahun 2021, Devi Yunita dan Ines Heidiani Ikasari melalui penelitian mereka yang membandingkan metode C4.5 dan Naïve Bayes dalam mengukur kepuasan pelanggan menunjukkan bahwa C4.5 lebih akurat dibandingkan dengan metode Naïve Bayes [3]. Penelitian ini mengungkapkan bahwa dalam konteks klasifikasi data mining, C4.5 memberikan hasil yang lebih presisi daripada Naïve Bayes.

Penelitian ini akan membandingkan algoritma Naïve Bayes dan C4.5. Hasil dari kedua algoritma tersebut akan dianalisis untuk menentukan mana yang memiliki nilai akurasi tertinggi sebagai metode terbaik untuk proses klasifikasi pada objek penelitian.

Recall dan *Precision* adalah dua metrik penting yang digunakan untuk mengevaluasi sistem klasifikasi, prediksi, atau pencarian. *Recall* adalah tingkat keberhasilan sistem

dalam menemukan kembali informasi yang relevan, sedangkan *Precision* adalah tingkat ketepatan informasi yang diberikan oleh sistem sesuai dengan yang diminta oleh pengguna.

II. METODOLOGI PENELITIAN

A. Naïve Bayes

Algoritma Naïve Bayes bekerja dengan menghitung probabilitas setiap kelas berdasarkan fitur yang ada, kemudian memilih kelas dengan probabilitas tertinggi sebagai prediksi [4]. Karena metode ini cepat dan memerlukan jumlah data yang relatif kecil untuk pelatihan, Naïve Bayes sering digunakan dalam situasi di mana waktu pemrosesan menjadi faktor penting.

Naïve Bayes juga dikenal karena kemampuannya untuk menangani data besar dan kompleks dengan efisiensi yang baik [5], meskipun asumsi independensinya terkadang dilanggar. Ini menjadikannya alat yang berguna dalam berbagai aplikasi, mulai dari prediksi risiko hingga analisis sentimen, dan banyak lagi.

$$P(H|X) = \frac{P(X|H) * P(H)}{P(X)} \quad \dots\dots\dots (1)$$

Keterangan:

- X = Data dengan class yang belum diketahui
- H = Hipotesis data merupakan suatu class spesifik
- P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)
- P(H) = Probabilitas hipotesis H (prior probabilitas)
- P(X|H) = Probabilitas X berdasarkan kondisi hipotesis H
- P(X) = Probabilitas X

Untuk menjelaskan metode Naïve Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode Naïve Bayes di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C) * P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad \dots\dots\dots (2)$$

Dimana variabel C mempresentasikan kelas, sementara variabel $F_1 \dots F_n$ mempresentasikan karakteristik petunjuk yang nanti dibutuhkan untuk melakukan klasifikasi. Maka dari itu rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*) dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{Prior * likelihood}{evidence} \quad \dots\dots\dots (3)$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan di bandingkan dengan nilai nilai posterior kelas lainnya untuk menentukan kelas apa suatu sampel akan diklasifikasikan.

B. Algoritma C4.5

Algoritma C4.5 merupakan salah satu teknik decision tree yang sering digunakan, yang menghasilkan beberapa aturan-aturan dan sebuah pohon keputusan dengan tujuan untuk meningkatkan keakuratan dari prediksi yang sedang dilakukan, disamping itu algoritma C4.5 merupakan algoritma yang mudah dimengerti .

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [6]:

- a. Pilih atribut sebagai akar.
- b. Buat cabang untuk tiap-tiap akar.
- c. Bagi kasus dalam cabang.
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Sebuah objek yang diklasifikasikan dalam pohon harus dites nilai entropinya [7]. Entropy adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impurity* dan *homogeneity* dari kumpulan data. Dari nilai entropy tersebut kemudian dihitung nilai *information gain* (IG) masing-masing atribut. Entropy (S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. Entropy dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai entropy maka akan semakin entropy digunakan dalam mengekstrak suatu kelas. Entropy digunakan untuk mengukur ketidakkaslian sistem informasi.

Besarnya entropy pada ruang sampel S didefinisikan dengan:

$$entropy(S) = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus} \dots\dots\dots (4)$$

Keterangan:

- S : Ruang (data) sampel digunakan untuk pelatihan
- p_{\oplus} : jumlah yang bersolusi positif atau mendukung pada data sampel untuk kriteria tertentu.
- p_{\ominus} : jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk kriteria tertentu.

Information gain adalah salah satu attribute selection measure yang digunakan untuk memilih test attribute tiap node pada tree. Atribut dengan informasi gain tertinggi dipilih sebagai test atribut dari suatu node.

Gain (S,A) merupakan perolehan informasi dari atribut A relative terhadap output data S. Perolehan informasi didapat dari output data atau variabel dependent S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (S,A).

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots (5)$$

Keterangan:

- S : himpunan kasus
- A : atribut
- N : jumlah partisi atribut A
- $|S_i|$: jumlah kasus pada partisi ke-i
- $|S|$: jumlah kasus pada S7

C. NSC Finance

PT. Nusa Surya Ciptadana (NSC Finance) adalah lembaga pembiayaan kendaraan bermotor yang merupakan bagian dari Nusantara Group. Didirikan pada tahun 2000 dan disahkan oleh Menteri Kehakiman Republik Indonesia pada 27 April 2001, NSC Finance kini memiliki cabang di 93 lokasi yang tersebar di berbagai provinsi di Indonesia.

Sejak Agustus 2004, NSC Finance mulai fokus membiayai mayoritas kredit di dealer resmi sepeda motor Honda, yaitu PT. Nusantara Sakti dan PT. Nusantara Surya Sakti. Kedua perusahaan ini merupakan "Sister Company" dari NSC Finance dan memiliki cabang di 9 kota di seluruh Indonesia. Dengan adanya kolaborasi ini, NSC Finance memiliki pangsa pasar tersendiri yang "captive" karena bergabung dengan jaringan dealer sepeda motor dalam satu atap. Hal ini sejalan dengan konsep Nusantara Group yang mengusung 3T: Tercepat, Termudah, dan Terpercaya dalam layanan kepemilikan kendaraan bermotor.

NSC Finance didukung oleh para direksi dari PT. Nusantara Sakti dan PT. Nusantara Surya Sakti yang telah berpengalaman dalam penjualan sepeda motor sejak tahun 1984. Sebagai bagian dari komitmen untuk memberikan layanan terbaik, NSC Finance juga menawarkan kesempatan pembiayaan kembali kepada nasabah potensial yang sebelumnya telah melakukan kredit, sehingga nasabah tetap dapat memperoleh pinjaman sesuai dengan kriteria yang ditentukan.

III. HASIL DAN PEMBAHASAN

A. Studi Kasus

Perusahaan yang bergerak di bidang dana modal usaha ini menawarkan pinjaman kepada nasabah potensial yang membutuhkan dana dengan memenuhi syarat-syarat tertentu. Penilaian terhadap calon nasabah dilakukan berdasarkan delapan atribut, yaitu umur, status pernikahan, pekerjaan, penghasilan, karakteristik, status pembayaran listrik, status pembayaran pinjaman, dan nilai pinjaman yang diajukan. Atribut-atribut ini digunakan untuk menilai apakah seorang nasabah layak menerima pinjaman lebih lanjut atau tidak.

Dua metode yang dapat digunakan untuk membantu sistem klasifikasi nasabah adalah Naïve Bayes dan C4.5, dengan pengukuran akurasi menggunakan metrik recall dan precision. Analisis perbandingan antara kedua metode ini bertujuan untuk mengetahui metode mana yang memberikan akurasi tertinggi dalam klasifikasi nasabah potensial. Dengan hasil perbandingan ini, perusahaan dapat memilih metode yang paling akurat dalam menentukan nasabah yang berhak melanjutkan pinjaman.

Hasil pengumpulan data di NSC Finance menghasilkan informasi data kriteria dan data training. Data kriteria yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Tabel Kriteria

ID	Kriteria	Sub Kriteria	Keterangan
1	Umur	DW = 21-35	Dewasa
		PB = 36-50	Paru Baya
		TU = 51-65	Tua
2	Status Pernikahan (s.Pernikahan)	BM	Belum Menikah
		MN	Menikah
		CR	Cerai
3	Pekerjaan	B	Berdagang
		KS	Karyawan Swasta
		PR	Profesi
		NL	Nelayan
		TR	Ternak
4	Penghasilan	BS	Besar
		SD	Sedang
		RD	Rendah
5	Karakteristik	BG	Bagus
		KR	Kurang
		TB	Tidak Bagus
6	Status Pembayaran Listrik (S.P. Listrik)	KC	Kecil
		S	Sedang
		T	Tinggi
		L	Lancar
7	Status Pembayaran (S. Pembayaran)	KL	Kurang Lancar
		TL	Tidak Lancar
8	Nilai Jaminan (N.J)	>2jt	>2jt
		>5jt-10jt	>5jt-10jt
		>11jt-20jt	>11jt-20jt

Data training dapat dilihat pada Tabel 2.

Tabel 2. Tabel Data Training

No	Nama	1	2	3	4	5	6	7	8	Pot
1	Latih1	PB	MN	KS	BS	BG	S	L	>2jt	B
2	Latih2	PB	MN	NL	BS	BG	KC	L	>5jt-10jt	B
3	Latih3	PB	MN	KS	BS	BG	T	L	>5jt-10jt	B
4	Latih4	DW	MN	KS	SD	BG	KC	KR	>5jt-10jt	DT
5	Latih5	DW	MN	B	SD	KR	KC	TL	>5jt-10jt	DT
6	Latih6	DW	MN	KS	BS	KR	S	TL	>5jt-10jt	TB
7	Latih7	PB	MN	B	SD	TB	KC	KL	>5jt-10jt	DT

8	Latih8	PB	MN	B	SD	BG	S	L	>5jt-10jt	B
9	Latih9	PB	MN	KS	BS	BG	S	L	>5jt-10jt	B
10	Latih10	DW	BM	KS	SD	BG	S	L	>5jt-10jt	B
...
31	Latih31	DW	MN	KS	SD	BG	S	L	>5jt-10jt	B
32	Latih32	PB	MN	TR	SD	BG	S	L	>2jt	B
33	Latih33	PB	MN	KS	SD	BG	S	KL	>5jt-10jt	TB
34	Latih34	TU	MN	B	RD	BG	S	KL	>5jt-10jt	TB
35	Latih35	PB	M	B	SD	BG	S	L	>5jt-10jt	B
36	Latih36	DW	BM	KS	SD	BG	S	L	>5jt-10jt	B
37	Latih37	PB	CR	KS	SD	BG	S	L	>2jt	B
38	Latih38	DW	MN	B	BS	BG	S	TL	>5jt-10jt	TB
39	Latih39	DW	MN	KS	SD	BG	S	TL	>5jt-10jt	TB
40	Latih40	DW	MN	KS	BS	BG	S	KL	>5jt-10jt	DT
...
61	Latih61	DW	MN	KS	BS	BG	S	L	>2jt	B
62	Latih62	PB	MN	B	RD	BG	KC	L	>5jt-10jt	B
63	Latih63	DW	MN	KS	SD	BG	S	L	>5jt-10jt	B
64	Latih64	DW	MN	B	SD	BG	S	L	>5jt-10jt	B
65	Latih65	DW	MN	TR	SD	BG	KC	L	>5jt-10jt	B
66	Latih66	PB	MN	B	BS	BG	T	L	>5jt-10jt	B
67	Latih67	TU	MN	B	BS	BG	S	L	>5jt-10jt	B
68	Latih68	PB	MN	KS	BS	BG	S	L	>5jt-10jt	B
69	Latih69	DW	BM	KS	SD	BG	S	KL	>5jt-10jt	DT
70	Latih70	PB	MN	B	SD	BG	T	TL	>5jt-10jt	TB
71	Latih71	DW	BM	KS	BS	BG	S	L	>5jt-10jt	B
72	Latih72	DW	BM	KS	BS	BG	S	L	>5jt-10jt	B
73	Latih73	TU	MN	B	SD	BG	KC	L	>5jt-10jt	TB
74	Latih74	TU	MN	B	SD	BG	S	L	>5jt-10jt	TB
75	Latih75	DW	BM	KS	BS	BG	S	L	>5jt-10jt	B
76	Latih76	TU	MN	B	SD	KR	S	KL	>5jt-10jt	DT
77	Latih77	TU	MN	PR	SD	TB	S	TL	>5jt-10jt	TB
78	Latih78	TU	MN	TR	SD	KR	S	TL	>5jt-10jt	TB
79	Latih79	TU	CR	B	SD	KR	S	TL	>5jt-10jt	TB
80	Latih80	PB	MN	KS	SD	BG	S	L	>5jt-10jt	B

Kolom potensi (pot) berisi hasil training yang mengklasifikasikan nasabah menjadi Berpotensi (B), Dipertimbangkan (DT) atau Tidak Berpotensi (TB).

1. Naïve Bayes

Tahapan awal yaitu menghitung nilai probabilitas masing-masing setiap kelas data training sebanyak 80 data dengan jumlah kelas sebanyak 3 yaitu berpotensi sebanyak 52, dipertimbangkan sebanyak 10, dan tidak berpotensi sebanyak 18.

- $P(\text{class}=\text{"Berpotensi"}) = 52/80 = 0,65$
- $P(\text{class}=\text{"Dipertimbangkan"}) = 10/80 = 0,125$
- $P(\text{class}=\text{"Tidak Berpotensi"}) = 18/80 = 0,225$

Selanjutnya menghitung setiap probabilitas atribut pada sub kriteria berdasarkan kelas potensial.

- $P(\text{umur} = \text{"Dewasa"} \mid \text{Potensial} = \text{"Berpotensi"}) = 26/52 = 0,50$
- $P(\text{umur} = \text{"Dewasa"} \mid \text{Potensial} = \text{"Dipertimbangkan"}) = 5/10 = 0,50$
- $P(\text{umur} = \text{"Dewasa"} \mid \text{Potensial} = \text{"Tidak Berpotensi"}) = 6/18 = 0,33$
- $P(\text{umur} = \text{"Paru Baya"} \mid \text{Potensial} = \text{"Berpotensi"}) = 24/52 = 0,46$
- $P(\text{umur} = \text{"Paru Baya"} \mid \text{Potensial} = \text{"Dipertimbangkan"}) = 2/10 = 0,20$
- $P(\text{umur} = \text{"Paru Baya"} \mid \text{Potensial} = \text{"Tidak Berpotensi"}) = 5/18 = 0,28$
- $P(\text{umur} = \text{"Tua"} \mid \text{Potensial} = \text{"Berpotensi"}) = 2/52 = 0,04$
- $P(\text{umur} = \text{"Tua"} \mid \text{Potensial} = \text{"Dipertimbangkan"}) = 3/10 = 0,30$
- $P(\text{umur} = \text{"Tua"} \mid \text{Potensial} = \text{"Tidak Berpotensi"}) = 7/18 = 0,39$

...

- $P(\text{Nilai Jaminan} = \text{">2jt"} \mid \text{Potensial} = \text{"Berpotensi"}) = 8/52 = 0,15$
- $P(\text{Nilai Jaminan} = \text{">2jt"} \mid \text{Potensial} = \text{"Dipertimbangkan"}) = 0/10 = 0,00$
- $P(\text{Nilai Jaminan} = \text{">2jt"} \mid \text{Potensial} = \text{"Tidak Berpotensi"}) = 0/18 = 0,00$
- $P(\text{Nilai Jaminan} = \text{">5jt-10jt"} \mid \text{Potensial} = \text{"Berpotensi"}) = 43/52 = 0,83$
- $P(\text{Nilai Jaminan} = \text{">5jt-10jt"} \mid \text{Potensial} = \text{"Dipertimbangkan"}) = 10/10 = 1,00$
- $P(\text{Nilai Jaminan} = \text{">5jt-10jt"} \mid \text{Potensial} = \text{"Tidak Berpotensi"}) = 18/18 = 1,00$
- $P(\text{Nilai Jaminan} = \text{">11jt-20jt"} \mid \text{Potensial} = \text{"Berpotensi"}) = 1/52 = 0,02$
- $P(\text{Nilai Jaminan} = \text{">11jt-20jt"} \mid \text{Potensial} = \text{"Dipertimbangkan"}) = 0/10 = 0,00$
- $P(\text{Nilai Jaminan} = \text{">11jt-20jt"} \mid \text{Potensial} = \text{"Tidak Berpotensi"}) = 0/18 = 0,00$

Berdasarkan perhitungan yang telah dilakukan, terdapat nilai "0" disetiap jumlah atribut. Namun dikarenakan basis perhitungannya bertumpu pada probabilitas tentu akan sangat beresiko, oleh sebab itu diperlukan teknik *Laplacian Correction* untuk menangani hal tersebut. Laplace Correction merupakan teknik yang digunakan untuk menyiasati supaya

probabilitas pada perhitungan naïve bayes tidak menghasilkan nilai 0 dikarenakan tidak adanya data untuk kategori tertentu dalam suatu class.

Persamaan dari Laplacian Correction dapat dilihat dari persamaan berikut:

$$P(c, t) = \frac{1+N(ct,D)}{|V|+N(c,D)} \dots\dots\dots (6)$$

Keterangan:

- $P(c,t)$: Probabilitas nilai yang ditemukan
- $N(ct,D)$: Jumlah Sample pada atribut P
- V : Jumlah nilai pada atribut
- $N(c,D)$: Jumlah sampel

Tabel 3. Probabilitas Atribut "Pekerjaan"

Pekerjaan	B	DT	TB	P(B)	P(DT)	P(TB)
B	17	4	10	0.33	0.4	0.56
KS	29	6	6	0.56	0.6	0.33
PR	2	0	1	0.04	0	0.06
NL	2	0	0	0.04	0	0
TR	2	0	1	0.04	0	0.06

Tabel 4. Perbaikan jumlah probabilitas atribut dengan Laplacian Correction

Pekerjaan	B	DT	TB	P(B)	P(DT)	P(TB)
B	17	5	11	0.33	0.50	0.61
KS	29	7	7	0.56	0.70	0.39
PR	2	1	2	0.04	0.10	0.11
NL	2	1	1	0.04	0.10	0.06
TR	2	1	2	0.04	0.10	0.11

Lakukan perhitungan tersebut dengan atribut yang memiliki nilai 0 sehingga memperoleh semua nilai probabilitasnya.

Selanjutnya melakukan perhitungan hasil variabel berdasarkan data testing. Diketahui data testing dapat dilihat pada Tabel 5.

Tabel 5. Data Testing

No	Nama	1	2	3	4	5	6	7	8	Pot
1	Latih1	PB	MN	B	SD	BG	S	L	>5jt-10jt	B
2	Latih2	TU	CR	TR	BS	BG	S	L	>5jt-10jt	TB
3	Latih3	PB	MN	B	RD	BG	K	KL	>2jt	DT
4	Latih4	PB	MN	TR	BS	BG	S	L	>5jt-10jt	B
5	Latih5	PB	MN	B	BS	BG	S	L	>5jt-10jt	B
6	Latih6	DW	MN	TR	BSS	BG	T	L	>11jt-20jt	B
7	Latih7	PB	MN	PR	SD	KR	S	TL	>5jt-10jt	TB
8	Latih8	DW	MN	KS	BS	BG	S	KL	>5jt-10jt	DT
9	Latih9	TU	CR	TR	BS	BG	S	L	>5jt-10jt	B

10	Latih10	PB	MN	PR	SD	TB	K	TL	>5jt-10jt	TB
11	Latih11	PB	MN	PR	SD	BG	S	L	>5jt-10jt	B
12	Latih12	PB	MN	B	SD	BG	K	L	>5jt-10jt	B
13	Latih13	PB	MN	PR	SD	BG	S	L	>5jt-10jt	B
14	Latih14	PB	MN	B	SD	BG	S	KL	>5jt-10jt	DT
15	Latih15	TU	MN	B	SD	KR	S	KL	>5jt-10jt	DT
16	Latih16	DW	MN	PR	SD	BG	K	L	>5jt-10jt	B
17	Latih17	TU	CR	B	SD	KR	S	TL	>5jt-10jt	TB
18	Latih18	DW	MN	B	BS	BG	S	KL	>5jt-10jt	DT
19	Latih19	DW	MN	KS	BS	BG	S	L	>5jt-10jt	B
20	Latih20	PB	MN	B	SD	BG	K	L	>5jt-10jt	B

Perhitungan hasil variabel berdasarkan data Latih1 :

- Untuk semua atribut potensial = “Berpotensi”
 $P(X | \text{Potensial} = \text{“Berpotensi”})$
 $= 0,46 \times 0,71 \times 0,33 \times 0,40 \times 1,00 \times 0,40 \times 1,02 \times 0,83$
 $= 0,0145991768$

Mengkalikan probability dengan semua atribut Potensial = “Berpotensi”
 $P(\text{Potensial} = \text{“Berpotensi”}) \times P(X | \text{Potensial} = \text{“Berpotensi”})$
 $= 0,65 \times 0,0145991768$
 $= 0,0094894649$

- Untuk semua atribut potensial = “Dipertimbangkan”
 $P(X | \text{Potensial} = \text{“Dipertimbangkan”})$
 $= 0,20 \times 0,70 \times 0,50 \times 0,90 \times 0,60 \times 0,90 \times 0,10 \times 1,10$
 $= 0,0037422$

Mengkalikan probability dengan semua atribut Potensial = “Dipertimbangkan”
 $P(\text{Potensial} = \text{“Dipertimbangkan”}) \times P(X | \text{Potensial} = \text{“Dipertimbangkan”})$
 $= 0,125 \times 0,0037422$
 $= 0,000467775$

- Untuk semua atribut potensial = “Tidak Berpotensi”
 $P(X | \text{Potensial} = \text{“Tidak Berpotensi”})$
 $= 0,28 \times 0,78 \times 0,61 \times 0,78 \times 0,61 \times 0,78 \times 0,17 \times 1,06$
 $= 0,0089095608$

Mengkalikan probability dengan semua atribut Potensial = “Tidak Berpotensi”
 $P(\text{Potensial} = \text{“Tidak Berpotensi”}) \times P(X | \text{Potensial} = \text{“Tidak Berpotensi”})$
 $= 0,225 \times 0,0089095608$
 $= 0,002000465118$

- Perbandingan Hasil Kelas
 $P(\text{Potensial} = \text{“Berpotensi”}) \times P(X | \text{Potensial} = \text{“Berpotensi”}) > P(\text{Potensial} = \text{“Dipertimbangkan”}) \times P(X | \text{Potensial} = \text{“Dipertimbangkan”})$

$P(\text{Potensial} = \text{“Berpotensi”}) \times P(X | \text{Potensial} = \text{“Berpotensi”}) > P(\text{Potensial} = \text{“Tidak Berpotensi”}) \times P(X | \text{Potensial} = \text{“Tidak Berpotensi”})$.

Untuk data latih 1 : Potensial = “Berpotensi” (perhitungan antara perkalian dari Potensial “Berpotensi” dengan Potensial “Dipertimbangkan” dengan Potensial “Tidak Berpotensi” menunjukkan bahwa nilai dari Potensial “Berpotensi” lebih besar dibandingkan dengan nilai Potensial “Dipertimbangkan” dan nilai Potensial “Tidak Berpotensi”.

Lakukan perhitungan tersebut untuk data latih yang lain sehingga memperoleh hasil klasifikasi potensial nasabah berdasarkan Naïve Bayes seperti Tabel 6.

Tabel 6. Hasil Klasifikasi Potensial Naïve Bayes

No	Nama	Potensial	Naïve Bayes
1	Latih1	B	B
2	Latih2	TB	TB
3	Latih3	DT	DT
4	Latih4	B	B
5	Latih5	B	B
6	Latih6	B	B
7	Latih7	TB	TB
8	Latih8	DT	DT
9	Latih9	B	TB
10	Latih 10	TB	TB
11	Latih11	B	B
12	Latih12	B	B
13	Latih13	B	B
14	Latih14	DT	DT
15	Latih15	DT	DT
16	Latih16	B	B
17	Latih17	TB	TB
18	Latih18	DT	DT
19	Latih19	B	B
20	Latih20	B	B

2. Algoritma C4.5

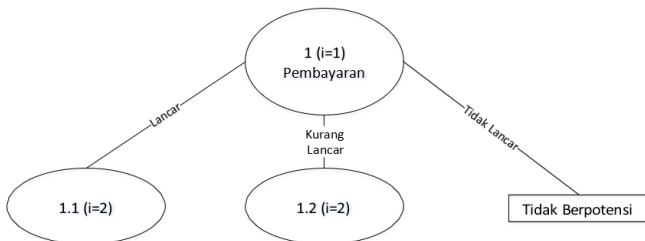
Dari data training yang ada pada Tabel 2 akan dilakukan proses klasifikasi menggunakan Algoritma C4.5. Langkah pertama yaitu memilih atribut sebagai akar yang didasari pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain, sebelumnya perlu dilakukan perhitungan jumlah kasus. Berikut hasil dari perhitungan ditunjukkan oleh Tabel 7.

Tabel 7. Perhitungan Entropy dan Gain (Node Akar)

Node	Atribut	Sub	Jumlah Kasus	B	DT	TB	Entropy	Gain
1	Umur		80	52	10	18	1.261	0.139
		DW	37	26	5	6	1.173	
		PB	31	24	2	5	0.966	
		TU	12	2	3	7	1.384	
	S.Nikah							0.074
		BM	17	14	2	1	0.834	
		MN	58	37	7	14	1.277	
		CR	5	1	1	3	1.371	
	Pekerjaan							0.073

B	29	15	4	10	1.416
KS	41	29	6	6	1.165
NL	2	2	0	0	0.000
PR	5	4	0	1	0.722
TR	3	2	0	1	0.918
Penghasilan					0.115
BS	34	29	2	3	0.745
SD	43	21	8	14	1.483
RD	3	2	0	1	0.918
Karakteristik					0.193
BG	68	51	6	11	1.045
KR	8	1	3	4	1.406
TB	4	0	1	3	0.811
S.P Listrik					0.020
T	6	5	0	1	0.650
S	62	40	8	14	1.274
KC	12	7	2	3	1.384
S. Bayar					0.956
L	55	52	0	3	0.305
KL	12	0	10	2	0.650
TL	13	0	0	13	0.000
N. Jaminan					0.084
>2	8	8	0	0	0.000
>5-10	71	43	10	18	1.347
>11-20	1	1	0	0	0.000

Berdasarkan perhitungan pada Tabel 7, dipilih atribut yang memiliki nilai gain tertinggi untuk dijadikan sebagai akar. Atribut yang dipilih yaitu Status Pembayaran sebagai akar. Terdapat tiga nilai dari atribut Status Pembayaran yaitu lancar, kurang lancar, dan tidak lancar. Dari ketiga nilai tersebut, nilai sub kriteria tidak lancar sudah mengklasifikasikan kasus menjadi 1 yaitu tidak berpotensi, sehingga tidak perlu dilakukan perhitungan lagi. Untuk nilai sub kriteria lancar dan kurang lancar masih perlu dilakukan perhitungan lagi. Dari hasil tersebut dapat digambarkan pohon keputusan sementara seperti pada Gambar 1.



Gambar 1. Pohon Keputusan Hasil Perhitungan Node 1

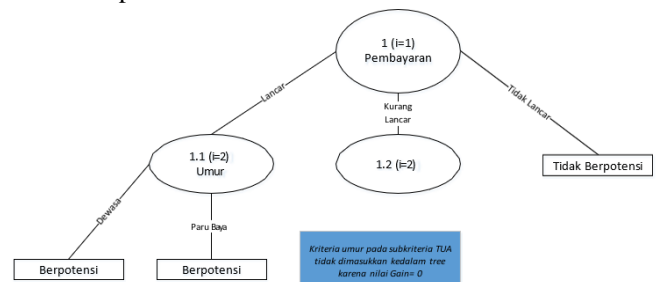
Selanjutnya menghitung nilai dari atribut yang lain untuk dijadikan node percabangan dari nilai atribut lancar. Perhitungan dilakukan dengan cara yang sama seperti pada Tabel 7 dan hasil perhitungan dapat dilihat pada Tabel 8.

Tabel 8. Perhitungan Entropy dan Gain (Node 1.1)

Node	Atribut	Sub	Jumlah Kasus	B	DT	TB	Entropy	Gain
1.1	S. Bayar	L	55	52	0	3	0.305	0.217
		Umur						
		DW	26	26	0	0	0.000	
		PB	24	24	0	0	0.000	
		TU	5	2	0	3	0.971	
	S. Nikah						0.026	
	BM	14	14	0	0	0.000		
	MN	40	37	0	3	0.384		

CR	1	1	0	0	0.000
Pekerjaan					0.093
B	18	15	0	3	0.650
KS	29	29	0	0	0.000
NL	2	2	0	0	0.000
PR	4	4	0	0	0.000
TR	2	2	0	0	0.000
Penghasilan					0.068
BS	29	29	0	0	0.000
SD	24	21	0	3	0.544
RD	2	2	0	0	0.000
Karakteristik					0.001
BG	54	51	0	3	0.310
KR	1	1	0	0	0.000
TB	0	0	0	0	0.000
S.P. Listrik					0.015
T	5	5	0	0	0.000
S	42	40	0	2	0.276
KC	8	7	0	1	0.544
N. Jaminan					0.014
>2	8	8	0	0	0.000
>5-10	46	43	0	3	0.348
>11-20	1	1	0	0	0.000

Berdasarkan perhitungan pada Tabel 8, dipilih atribut Umur sebagai node percabangan dari nilai sub kriteria lancar pada Tabel 8. Terlihat ketiga nilai tersebut, nilai sub kriteria dewasa sudah mengklasifikasikan kasus menjadi 1, yaitu keputusannya berpotensi. Kemudian nilai sub kriteria paru baya sudah mengklasifikasikan kasus menjadi 1, yaitu berpotensi sehingga tidak perlu dilakukan perhitungan lagi. Dari hasil tersebut dapat digambarkan pohon keputusan sementara pada Gambar 2.



Gambar 2. Pohon Keputusan Hasil Perhitungan Node 1.1

Selanjutnya menghitung nilai atribut penghasilan untuk dijadikan node percabangan dari nilai subkriteria kurang lancar. Perhitungan dilakukan dengan cara yang sama pada Tabel 8 dan hasil perhitungan dapat dilihat pada Tabel 9.

Tabel 9. Perhitungan Entropy dan Gain (Node 1.2)

Node	Atribut	Sub	Jumlah Kasus	B	DT	TB	Entropy	Gain
1.2	S. Bayar	KL	12	0	10	2	0.650	0.150
		Umur						
		DW	5	0	5	0	0.000	
		PB	3	0	2	1	0.918	
		TU	4	0	3	1	0.811	
	S. Nikah						0.077	
	BM	2	0	2	0	0.000		
	MN	9	0	7	2	0.764		
	CR	1	0	1	0	0.000		

Pekerjaan						0.073
B	29	15	4	10	1.416	
KS	41	29	6	6	1.165	
NL	2	2	0	0	0.000	
PR	5	4	0	1	0.722	
TR	3	2	0	1	0.918	
Penghasilan						0.273
BS	2	0	2	0	0.000	
SD	9	0	8	1	0.503	
RD	1	0	0	1	0.000	
Karakteristik						0.109
BG	8	0	6	2	0.811	
KR	3	0	3	0	0.000	
TB	1	0	1	0	0.000	
S.P. Listrik						0.048
T	0	0	0	0	0.000	
S	10	0	8	2	0.722	
KC	2	0	2	0	0.000	
N.Jaminan						0.000
>2	0	0	0	0	0.000	
>5-10	12	0	10	2	0.650	
>11-20	0	0	0	0	0.000	

Gambar 4. Pohon Keputusan Hasil Perhitungan Node 1.2.1

Dari pembuatan pohon keputusan yang sudah terbentuk maka diperoleh aturan atau rule.

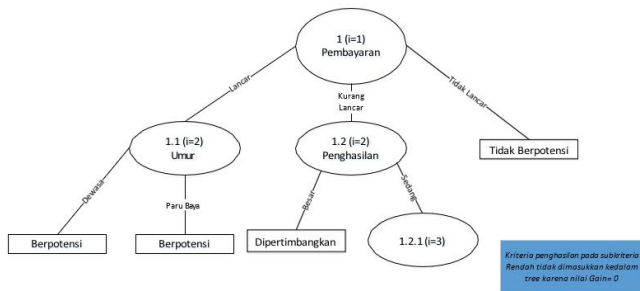
- Jika status pembayaran tidak lancar maka tidak berpotensi.
- Jika status pembayaran lancar dan usia dewasa maka berpotensi.
- Jika status pembayaran lancar dan usia paru baya maka berpotensi.
- Jika status pembayaran kurang lancar dan penghasilan besar maka dipertimbangkan.
- Jika status pembayaran kurang lancar dan penghasilan rendah dan usia dewasa maka dipertimbangkan.

Rule yang telah diperoleh lakukan proses klasifikasi terhadap data testing yang ada pada Tabel 5 sehingga diperoleh hasil klasifikasi potensi nasabah berdasarkan Algoritma C4.5.

Tabel 10. Hasil Klasifikasi Algoritma C4.5

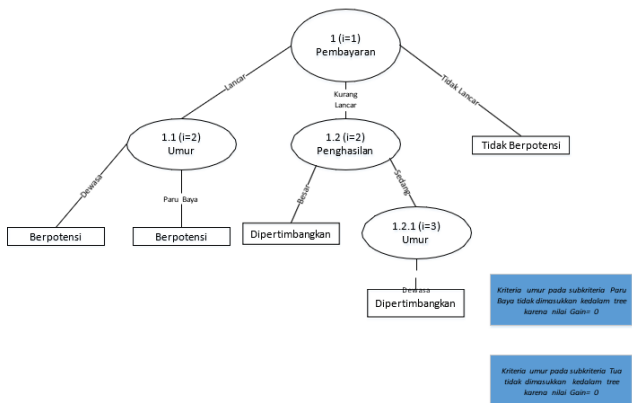
No	Nama	Potensial	C4.5
1	Latih1	B	B
2	Latih2	TB	B
3	Latih3	DT	B
4	Latih4	B	B
5	Latih5	B	B
6	Latih6	B	B
7	Latih7	TB	TB
8	Latih8	DT	DT
9	Latih9	B	DT
10	Latih 10	TB	TB
11	Latih11	B	B
12	Latih12	B	B
13	Latih13	B	B
14	Latih14	DT	DT
15	Latih15	DT	DT
16	Latih16	B	B
17	Latih17	TB	TB
18	Latih18	DT	DT
19	Latih19	B	B
20	Latih20	B	B

Berdasarkan perhitungan pada Tabel 9, dipilih atribut penghasilan sebagai node percabangan dari nilai subkriteria kurang lancar. Pada Tabel 9 terlihat ketiga nilai tersebut, nilai subkriteria besar sudah mengklasifikasikan kasus menjadi 1, yaitu dipertimbangkan. Untuk nilai subkriteria sedang masih diperlukan nilai perhitungan lagi. Dari hasil tersebut dapat digambarkan pohon keputusan sementara seperti pada Gambar 3.



Gambar 3. Pohon Keputusan Hasil Perhitungan Node 1.2

Lakukan perhitungan tersebut hingga terbentuk pohon keputusan terakhir. Diketahui percabangan berakhir di node 1.2.1.



3. Perbandingan Akurasi

Salah satu aspek yang menjadi parameter keandalan dari suatu algoritma klasifikasi adalah akurasi. Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual. Precision adalah tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan oleh sistem. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali informasi. Berikut perhitungan akurasi, precision, dan recall:

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + FP + TN} * 100\%$$

$$\text{Precision} = \frac{TP}{TP + FP} * 100\%$$

$$\text{Recall} = \frac{TP}{TP + FN} * 100\%$$

Keterangan:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

Pada Tabel 6 telah diperoleh hasil klasifikasi potensi nasabah menggunakan metode Naïve Bayes dan Tabel 10 telah diperoleh hasil klasifikasi potensi nasabah menggunakan metode C4.5. Dari hasil tersebut akan dilakukan proses perbandingan akurasi berdasarkan potensi nasabah pada data testing dan potensi nasabah berdasarkan hasil klasifikasi metode. Akurasi hasil klasifikasi metode Naïve Bayes dan C4.5 dapat dilihat pada Tabel 11.

Tabel 11. Perbandingan Akurasi Metode

No	Nama	Akurasi Naïve Bayes			Akurasi C4.5		
		Pot.	Naïve Bayes	Analisa	Pot.	C4.5	Analisa
1	Latih1	B	B	TP	B	B	TP
2	Latih2	TB	TB	TP	TB	B	FP
3	Latih3	DT	DT	TP	DT	B	FP
4	Latih4	B	B	TP	B	B	TP
5	Latih5	B	B	TP	B	B	TP
6	Latih6	B	B	TP	B	B	TP
7	Latih7	TB	TB	TP	TB	TB	TP
8	Latih8	DT	DT	TP	DT	DT	TP
9	Latih9	B	TB	FN	B	DT	TP
10	Latih10	TB	TB	TP	TB	TB	TP
11	Latih11	B	B	TP	B	B	TP
12	Latih12	B	B	TP	B	B	TP
13	Latih13	B	B	FP	B	B	FP
14	Latih14	DT	DT	TP	DT	DT	TP
15	Latih15	DT	DT	FN	DT	DT	FN
16	Latih16	B	B	TP	B	B	TP
17	Latih17	TB	TB	TP	TB	TB	TP
18	Latih18	DT	DT	TP	DT	DT	TP
19	Latih19	B	B	TP	B	B	TP
20	Latih20	B	B	TP	B	B	TP

Naïve Bayes

Diketahui jumlah TP = 17, TN = 0, FP = 1 dan FN = 2, Maka perhitungan akurasi, precision, dan recall berdasarkan untuk hasil metode Naïve Bayes :

$$\text{Akurasi} = \frac{17 + 0}{17 + 2 + 1 + 0} * 100\% = 85,00\%$$

$$\text{Precision} = \frac{17}{17 + 1} * 100\% = 94,44\%$$

$$\text{Recall} = \frac{17}{17 + 2} * 100\% = 89,47\%$$

C4.5

Diketahui jumlah TP = 15, TN = 0, FP = 3 dan FN = 2, Maka perhitungan akurasi, precision, dan recall berdasarkan untuk hasil metode C4.5 :

$$\text{Akurasi} = \frac{15 + 0}{15 + 2 + 3 + 0} * 100\% = 75,00\%$$

$$\text{Precision} = \frac{15}{15 + 3} * 100\% = 83,33\%$$

$$\text{Recall} = \frac{15}{15 + 2} * 100\% = 88,24\%$$

Setelah dilakukan pengujian terhadap 20 data testing dengan metode Naïve Bayes dan metode C4.5 didapatkan hasil bahwa akurasi Naïve Bayes lebih besar 85.00% dibandingkan dengan akurasi C4.5 75.00%.

B. Uji Coba

Untuk menunjukkan bagaimana metode Naïve Bayes dan C4.5 dapat bekerja, maka penulis membangun sebuah program berbasis desktop.

Id. Client	Nama	Keputusan	Prediksi Naive Bayes	Analisa
KTS01	Mustafa	Berpotensi	Berpotensi	TP
KTS02	Menneng	Tidak Berpotensi	Tidak Berpotensi	TP
KTS03	Amahyulaka	Diperimbangkan	Diperimbangkan	TP
KTS04	Syahri	Berpotensi	Berpotensi	TP
KTS05	Muhammad Rozi	Berpotensi	Berpotensi	TP
KTS06	Heis Wandi	Berpotensi	Berpotensi	TP
KTS07	Muhammad Thoyib	Tidak Berpotensi	Tidak Berpotensi	TP
KTS08	Damawan Diputra	Diperimbangkan	Diperimbangkan	TP
KTS09	Pacinongi	Berpotensi	Diperimbangkan	FN
KTS10	Alasabit	Tidak Berpotensi	Tidak Berpotensi	TP
KTS11	Roman Prayuda	Berpotensi	Berpotensi	TP
KTS12	Anastasia Pallunan	Berpotensi	Berpotensi	TP
KTS13	Rahmat Budiono	Tidak Berpotensi	Berpotensi	FP
KTS14	Agustinah	Diperimbangkan	Diperimbangkan	TP
KTS15	Lorensius Ontoni	Berpotensi	Diperimbangkan	FN
KTS16	Ludgerus Bungsu Say	Berpotensi	Berpotensi	TP
KTS17	Rosta	Tidak Berpotensi	Tidak Berpotensi	TP
KTS18	Selviyani	Diperimbangkan	Diperimbangkan	TP
KTS19	ABD Aziz	Berpotensi	Berpotensi	TP
KTS20	Sri Astilaningah	Berpotensi	Berpotensi	TP

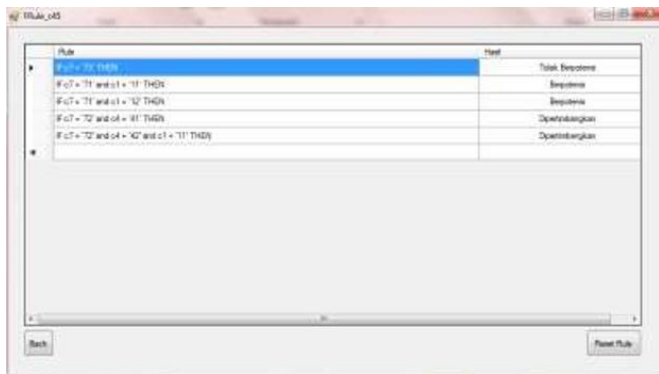
True Positif (TP) 17 Accuracy 85.00%
 True Negative (TN) 0 Precision 94.44%
 False Positive (FP) 1 Recall 89.47%
 False Negative (FN) 2

Gambar 5. Hasil Akurasi Metode Naïve Bayes

Id. Client	Nama	Keputusan	Prediksi C4.5	Analisa
1001	Mustafa	Berpotensi	Berpotensi	TP
1002	Menneng	Tidak Berpotensi	Berpotensi	FP
1003	Amahyulaka	Diperimbangkan	Berpotensi	FP
1004	Syahri	Berpotensi	Berpotensi	TP
1005	Muhammad Rozi	Berpotensi	Berpotensi	TP
1006	Heis Wandi	Berpotensi	Berpotensi	TP
1007	Muhammad Thoyib	Tidak Berpotensi	Tidak Berpotensi	TP
1008	Damawan Diputra	Diperimbangkan	Diperimbangkan	TP
1009	Pacinongi	Berpotensi	Diperimbangkan	FN
1010	Alasabit	Tidak Berpotensi	Tidak Berpotensi	TP
1011	Roman Prayuda	Berpotensi	Berpotensi	TP
1012	Anastasia Pallunan	Berpotensi	Berpotensi	TP
1013	Rahmat Budiono	Tidak Berpotensi	Berpotensi	FP
1014	Agustinah	Diperimbangkan	Diperimbangkan	TP
1015	Lorensius Ontoni	Berpotensi	Diperimbangkan	FN
1016	Ludgerus Bungsu Say	Berpotensi	Berpotensi	TP
1017	Rosta	Tidak Berpotensi	Tidak Berpotensi	TP
1018	Selviyani	Diperimbangkan	Diperimbangkan	TP
1019	ABD Aziz	Berpotensi	Berpotensi	TP
1020	Sri Astilaningah	Berpotensi	Berpotensi	TP

True Positif (TP) 15 Accuracy 75.00%
 True Negative (TN) 0 Precision 83.33%
 False Positive (FP) 3 Recall 88.24%
 False Negative (FN) 2

Gambar 6. Hasil Akurasi Metode C4.5



Gambar 7. Rule Metode C4.5

Gambar 5 menunjukkan akurasi model Naïve Bayes dalam mengelompokkan data. Model ini bekerja dengan baik jika fitur-fitur dianggap tidak saling bergantung. Walaupun hasil akurasi, precision, dan recall sudah bagus, tetap perlu diperhatikan bahwa distribusi data dan hubungan antar fitur bisa memengaruhi hasilnya.

Gambar 6 menunjukkan hasil akurasi model klasifikasi dengan metode C4.5. Metode ini menggunakan pohon keputusan untuk memisahkan data berdasarkan fitur-fitur tertentu. Akurasi yang ditampilkan menggambarkan seberapa baik model ini memprediksi data dengan benar. Metode C4.5 cukup efektif dalam menangani data yang beragam, namun penting untuk memperhatikan bahwa jika pohon keputusan terlalu rumit, model bisa jadi tidak bekerja baik pada data baru.

IV. KESIMPULAN

Dari hasil penelitian, klasifikasi dilakukan dengan menggunakan 80 data training dan 20 data testing. Hal ini menyebabkan kemungkinan ada beberapa data yang hasil klasifikasinya tidak sesuai dengan data aslinya. Dalam perbandingan antara metode Naïve Bayes dan C4.5, Naïve Bayes terbukti lebih akurat. Naïve Bayes memiliki akurasi 85%, precision 94.44%, dan recall 89.47%, yang lebih baik dibandingkan dengan metode C4.5.

REFERENSI

- [1] K. Indriani and Q. Tanjung, "Sistem Pendukung Keputusan Kelayakan Kredit Motor Menggunakan Metode NAÏVE BAYES Pada NSC FINANCE Cikampek."
- [2] L. Navia *et al.*, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," vol. 1, no. 2, 2016.
- [3] D. Yunita and I. H. Ikasari, "Perbandingan Metode Klasifikasi C4.5 dan Naive Bayes untuk Mengukur Kepuasan Pelanggan," vol. 6, no. 3, pp. 2622–4615, 2021, doi: 10.32493/informatika.v6i3.9160.
- [4] H. Susana and N. Suarna, "PENERAPAN MODEL KLASIFIKASI METODE NAIVE BAYES TERHADAP PENGGUNAAN AKSES INTERNET Program Studi Teknik Informatika STMIK IKMI Cirebon Jl Perjuangan No 10B Kesambi Kota Cirebon 3) Program Studi Rekayasa Perangkat Lunak STMIK IKMI Cirebon Jl Perjuangan No 10B Kesambi Kota Cirebon 4) Program Studi Komputerisasi Akuntansi STMIK IKMI Cirebon Jl Perjuangan No 10B Kesambi Kota Cirebon," *Jurnal Sistem Informasi dan Teknologi Informasi*, vol. 4, no. 1, pp. 1–8, 2022.
- [5] H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *Jurnal Teknologi Informasi dan Komunikasi (TIKOMSiN)*, vol. 8, no. 2, Oct. 2020, doi: 10.30646/tikomsin.v8i2.500.
- [6] N. Azwanti, "ALGORITMA C4.5 UNTUK MEMPREDIKSI MAHASISWA YANG MENGULANG MATA KULIAH (STUDI KASUS DI AMIK LABUHAN BATU)," *Jurnal SIMETRIS*, vol. 9, no. 1, 2018.
- [7] L. Bachtiar and M. Mahradianur, "Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai," *Jurnal Informatika*, vol. 10, no. 1, pp. 28–36, Mar. 2023, doi: 10.31294/inf.v10i1.15115.