

# PERBANDINGAN ALGORITMA C4.5 DAN NAIVE BAYES UNTUK MENGUKUR MINAT PENJUALAN SEPATU

<sup>1</sup>Nur Azizah Lubis, <sup>1</sup>M. Safii, <sup>2</sup>Ommi Alfina✉

<sup>1</sup>Program Studi Sistem Informasi STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

<sup>2</sup>Universitas Potensi Utama, Medan, Indonesia

Email: [ny.aeroen@gmail.com](mailto:ny.aeroen@gmail.com)

DOI: <https://doi.org/10.46880/methoda.Vol13No3.pp337-345>

## ABSTRACT

*The Second Gangbrand shoe shop is one that sells second-hand shoes or what you could also call quality and original second-hand shoes that have certain brands at affordable prices that are cheaper than the original price. This research aims to measure the level of customer interest by comparing the C4.5 algorithm method and the Naive Bayes algorithm. The data source was obtained from second gangbrand stores which were taken based on customer interest. So it is necessary to carry out data analysis to classify customer interest data using the C4.5 and Naive Bayes algorithms to compare accuracy and precision which are the benchmarks in this research. Calculations in this research were carried out manually using Microsoft Excel according to the C4.5 and Naive Bayes algorithm calculation models and then evaluated using the Rapidminer 10.3 tool which was used to help determine accurate values. After conducting research testing, the C4.5 algorithm received an accuracy value of 60.00% and a precision of 50.00%, while the Naive Bayes algorithm received an accuracy value of 60% and a precision of 33.33%. So it can be concluded that the two algorithms have the same accurate accuracy value, but in terms of precision value the C4.5 algorithm is superior in determining customer interest recommendations. It is hoped that the results of this research can provide input and information for future researchers.*

**Keyword:** C4.5 Algorithm, Naive Bayes Algorithm, Classification, Accuracy, Rapidminer.

## ABSTRAK

*Toko sepatu SecondGangbrand merupakan salah satu penjualan sepatu second atau bisa disebut juga sepatu bekas yang berkualitas dan ori yang mempunyai merek tertentu dengan harga yang terjangkau lebih murah daripada harga aslinya. Penelitian ini bertujuan untuk mengukur tingkat minat pelanggan dengan perbandingan metode algoritma C4.5 dan algoritma Naive Bayes. Sumber data yang diperoleh dari toko second gangbrand yang diambil berdasarkan minat pelanggan. Maka perlu dilakukan adanya analisa data untuk mengklasifikasi data minat pelanggan menggunakan algoritma C4.5 dan Naive Bayes untuk membandingkan akurasi dan presisi dimana menjadi tolak ukur dalam penelitian ini. Perhitungan pada penelitian ini dilakukan dengan cara manual menggunakan Microsoft excel sesuai dengan model perhitungan algoritma C4.5 dan Naive Bayes kemudian dievaluasi dengan menggunakan tools rapidminer 10.3 yang digunakan untuk membantu menentukan nilai yang akurat. Setelah dilakukan pengujian penelitian tersebut algoritma C4.5 mendapat nilai akurasi sebesar 60.00% dan presisi 50.00% sedangkan algoritma naive bayes mendapat nilai akurasi sebesar 60% dan presisi 33,33%. Sehingga dapat disimpulkan kedua algoritma tersebut memiliki nilai akurasi yang sama akurat, namun pada segi nilai presisi algoritma C4.5 lebih unggul dalam penentuan rekomendasi*

*minat pelanggan. Diharapkan hasil penelitian ini bisa menjadi masukan dan informasi bagi peneliti selanjutnya.*

**Kata Kunci:** *Algoritma C4.5, Algoritma Naive Bayes, Klasifikasi, Akurasi, Rapidminer.*

---

## **PENDAHULUAN**

Data mining atau penggalian data pada era digital telah menjadi langkah strategis untuk digunakan pada berbagai bidang. Salah satu bidang yang menggunakan data mining adalah bidang pemasaran. Di bidang pemasaran, data mining dapat memberikan gambaran tingkat minat suatu barang dan membantu penyusunan strategi pemasaran. Selain itu, keterbatasan kemampuan manusia untuk melakukan perhitungan dan pengolahan data secara manual tanpa bantuan komputer membuat perolehan informasi juga terbatas. Oleh karena itu, penerapan data mining untuk klasifikasi tingkat minat dari suatu barang sangat diperlukan.

Pada saat ini fashion yang ada di Indonesia semakin berkembang dengan mengikuti arus modernisasi. Dengan perkembangan yang ada ini menjadikan masyarakat sebagai konsumen yang cukup selektif dalam memilih gaya hidupnya. Produk fashion sepatu yang bermerek merupakan salah satu kategori produk yang sering dibeli oleh masyarakat secara online maupun offline. Minat masyarakat pada produk sepatu second yang bermerek dan berkualitas pun meningkat untuk memnuhi gaya hidup seseorang. Pembelian sepatu second mengalami perkembangan terus menerus dalam setiap hari nya pada bidang perdagangan di era modern saat ini banyak terjadi perkembangan pesat dimana-mana bukan hanya terjadi pada pasar-pasar tradisional saja melainkan menggunakan jasa online yang sangat mempermudah pelanggan dalam pembelian produk sepatu second untuk keperluan pribadi maupun keperluan keluarga. Pada minat masyarakat yang tinggi untuk membeli sepatu second, akan tetapi ada beberapa faktor yang menjadi perhitungan masyarakat untuk membeli dengan harus memilih kualitas sepatu yang baik dan terjangkau dengan harga yang terjangkau.

Berdasarkan uraian diatas maka penulis melakukan analisa dari kumpulan data toko sepatu secong gangbrand. Adapun dipilihnya

Metode algoritma Naive Bayes dan metode C4.5 tujuannya sebagai algoritma pengklasifikasi minat kepuasan pelanggan karena tingkat akurasi dari kedua algoritma ini relatif tinggi. Untuk klasifikasi rekomendasi pada minat produk sepatu second berdasarkan atributnya untuk menghasilkan perbandingan klasifikasi akurasi dan presisi dari algoritma C4.5 (Decision Tree) dan algoritma Naive Bayes untuk diproses dalam Data Mining. Data yang telah di dapatkan selanjutnya akan dihitung manual menggunakan Microsoft excel dan akan di uji menggunakan RapidMiner 10.3.

## **STUDI LITERATUR**

### **Data Mining**

Data mining adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar. Proses data mining seringkali menggunakan metode statistika, matematika, hingga memanfaatkan teknologi artificial intelligence. Nama alternatifnya yaitu Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, dan lain-lain. Data mining memiliki banyak sekali fungsi, Untuk fungsi utamanya sendiri yaitu ada dua, yaitu fungsi descriptive dan fungsi predictive (Amalia, 2020):

#### 1) Descriptive

Fungsi deskripsi dalam data mining adalah sebuah fungsi untuk memahami lebih jauh tentang data yang di amati. Dengan melakukan sebuah proses di harap bisa mengetahui perilaku dari sebuah data tersebut. Data tersebut itulah yang nantinya dapat di gunakan untuk mengetahui karakteristik dari data yang dimaksud.

#### 2) Predictive

Fungsi prediksi merupakan sebuah fungsi bagaimana sebuah proses nantinya akan menemukan pola tertentu dari suatu data.

Pola-pola tersebut dapat di ketahui dari berbagai variabel-variabel yang ada pada data.

Adapun proses atau tahapan-tahapan di dalam proses data mining (Aprilianur & Hadisaputro, 2022) diantaranya:

- Data Cleansing, Proses di mana data-data yang tidak lengkap, mengandung error dan tidak konsisten di buang dari koleksi data. Ketahui juga data lifecycle management untuk mengetahui tentang pengolahan data.
- Data Integration, Proses integrasi data di mana yang berulang akan di kombinasikan.
- Selection, Proses seleksi atau pemilihan data yang relevan terhadap analisis untuk di terima dari koleksi data yang ada.
- Data Transformation, Proses transformasi data yang sudah di pilih ke dalam bentuk mining procedure melalui cara dan agresi data.
- Data Mining, Proses yang paling penting dimana akan dilakukan berbagai teknik yang di aplikasikan untuk mengekstrak berbagai pola-pola potensial untuk mendapatkan data yang berguna.
- Pattern Evolution, Sebuah proses di mana pola-pola menarik yang sebelumnya sudah di temukan dengan identifikasi berdasarkan measure yang telah di berikan
- Knowledge Presentation, Merupakan proses tahap terakhir, Dalam hal ini di gunakan teknik visualisasi yang bertujuan membantu user dalam mengerti dan menginterpretasikan hasil dari penambahan data.

#### Algoritma C4.5

Algoritma C4.5 merupakan ekstensi dari algoritma ID3 dan menggunakan prinsip Decision Tree yang mirip. Algoritma C4.5 mempunyai kelebihan dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain. Langkah dalam pemrosesan data menggunakan algoritma C4.5 ada 4 tahapan, yaitu mencari nilai entropy, mencari nilai Gain, pembentukan pohon keputusan dan rule (Salim & Azis, 2018).

#### 1. Entropy

$$\text{Entropy}(S) = \sum_{i=0}^n - p_i \log_2 (p_i)$$

Dimana :

S : Himpunan kasus

n : Jumlah partisi S

pi : Jumlah kasus pada partisi ke-i

#### 2. Mencari nilai Gain

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

Dimana :

S : Himpunan kasus

n : Jumlah partisi

Si : Jumlah kasus pada kasus ke-i

S : Jumlah kasus dalam S

#### 3. Pembuatan pohon keputusan

#### 4. Membuat role berdasarkan pohon keputusan.

#### 5. Diketahui model klasifikasinya

### Algoritma Naïve Bayes

Naïve Bayes adalah algoritma klasifikasi yang mudah, sederhana dan sering digunakan untuk klasikasi dokumen. Selain itu Naïve Bayes juga digunakan untuk pengambilan keputusan berdasarkan hasil klasifikasi yang diperoleh (Peng et al., 2018). Dalam pemrosesan kedalam algoritma Naïve Bayes ada beberapa tahapan, yaitu:

1. Penentuan nilai Prior atau nilai probabilitas.
2. Menghitung nilai probabilitas tiap kelas
3. Membandingkan hasil kelas PC1 dan PC2
4. Diketahui model klasifikasinya.

### Rapidminer Studio

RapidMiner ditulis dalam bahasa pemrograman Java. RapidMiner menyediakan GUI untuk merancang dan menjalankan alur kerja analitik. Alur kerja tersebut disebut "Proses" di RapidMiner dan terdiri dari beberapa "Operator". Setiap operator melakukan satu tugas dalam proses, dan output dari masing-masing operator menjadi input dari operator berikutnya. Alternatifnya, mesin dapat dipanggil dari program lain atau digunakan sebagai API. Fungsi individu dapat dipanggil dari baris perintah (Ainurrohmah, 2021). RapidMiner menyediakan skema pembelajaran, model, dan algoritme dan dapat diperluas menggunakan skrip R dan Python. Menurut Bloor Research, RapidMiner menyediakan 99% solusi analitik

tingkat lanjut melalui kerangka kerja berbasis template yang mempercepat pengiriman dan mengurangi kesalahan hamper menghilangkan kebutuhan untuk menulis kode Rapidminer menyediakan prosedur penambahan data dan pembelajaran mesin termasuk : pemuatan dan transformasi data (ETL), prapemrosesan dan visualisasi data, analitik prediktif dan pemodelan statistik, evaluasi, dan penerapan (Ainurrohmah, 2021).

## METODE PENELITIAN

### Pengumpulan Data Set

Pada penelitian ini metode pengumpulan data yang digunakan adalah studi dokumen. Studi dokumen merupakan Pengambilan data melalui dokumen tertulis maupun elektronik dari toko penjualan sepatu gangbrand.

### Seleksi Data

Seleksi Data merupakan proses pemilihan data yang akan digunakan dalam proses data mining.

**Tabel 1.** Data Hasil Pra-Pengolahan

No	Kode Sepatu	Produk	Kualitas	Label	Harga	Status
1	001ABC	Palsu	Bagus	Ada	Rendah	Minat
2	002ABC	Original	Tidak Bagus	Ada	Sedang	Minat
3	003ABC	Original	Bagus	Ada	Tinggi	Tidak Minat
4	004ABC	Original	Bagus	Ada	Sedang	Tidak Minat
5	005ABC	Original	Tidak Bagus	Ada	Rendah	Minat
6	006ABC	Original	Bagus	Ada	Sedang	Tidak Minat
7	007ABC	Original	Tidak Bagus	Ada	Rendah	Minat
8	008ABC	Original	Tidak Bagus	Ada	Sedang	Minat
9	0009BC	Original	Bagus	Ada	Sedang	Minat
10	0010ABC	Original	Bagus	Ada	Tinggi	Tidak Minat
11	0011ABC	Palsu	Bagus	Ada	Rendah	Tidak Minat
12	0012ABC	Original	Bagus	Tidak Ada	Rendah	Minat
13	0013ABC	Palsu	Bagus	Ada	Rendah	Minat
14	0014ABC	Original	Bagus	Ada	Tinggi	Minat
15	0015ABC	Palsu	Bagus	Ada	Tinggi	Tidak Minat
16	0016ABC	Original	Bagus	Ada	Sedang	Minat
17	0017ABC	Original	Tidak Bagus	Ada	Rendah	Minat
18	0018ABC	Original	Bagus	Tidak Ada	Rendah	Minat
19	0019ABC	Original	Bagus	Ada	Rendah	Tidak Minat
20	0020ABC	Original	Tidak Bagus	Ada	Rendah	Minat
21	0021ABC	Original	Bagus	Ada	Tinggi	Tidak Minat
22	0022ABC	Original	Bagus	Ada	Sedang	Tidak Minat

### Gambaran Umum Penelitian

Informasi primer yang dipakai didalam metode penelitian ini diperoleh dari data pada toko sepatu second gangbrand. Penelitian ini berhubungan dengan masalah yang terkait dengan minat pelanggan. Setelah data diperoleh, lalu tentukan variabel-variabelnya yang akan dipakai. Variabel respon diambil dari minat dan

tidak minat pelanggan pada sepatu second gangbrand. Sedangkan variabel bebasnya, yaitu produk, kualitas, label, harga, status keterangan. Tahapan-tahapan penelitian yang dilakukan antara lain:

1. Pra-pengolahan data.
2. Hitung manual data tersebut untuk menentukan jumlah minat dan tidak minat

3. Hasil perhitungannya diolah ke dalam perhitungan dengan metode algoritma Naïve Bayes dan Metode algoritma C4.5 mengupayakan akurasi analisis data dengan data uji.
4. Peneliti menggunakan aplikasi RapidMiner untuk menguji akurasi dari metode Naïve Bayes dan metode C4.5.
5. Hasil perbandingan algoritmanya dapat digunakan untuk mengetahui akurasi terbaik terhadap minat pelanggan. Penelitian ini menerapkan algoritma dengan 2 jenis klasifikasi, yaitu klasifikasi C4.5 dan klasifikasi Naïve Bayes.

Langkah-langkah penerapan analisis metode klasifikasi dalam penelitian ini adalah (Mardi, 2017):

- a. Algoritma C4.5 :
  1. Hitung jumlah respon yang puas, jumlah respon yang tidak puas, serta entropy dari keseluruhan.
  2. Hitung gain untuk setiap atribut.
  2. Tentukan node root untuk mendapatkan nilai gain tertinggi.
  3. Lalu hitung jumlah kasus, jumlah respon yang puas, jumlah respon tidak puas, dan entropy atas seluruh permasalahan dan permasalahan yang dibagi dengan tanda selain node root, lalu gunakan node root rendah nilai atributnya, sehingga seluruh

atribut dapat menjadi node dan membentuk pohon keputusan.

- b. Algoritma Naïve Bayes (Susanti et al., 2022)
  1. Data latih diinput.
  2. Hitung jumlah data, probabilitas, dan bila datanya numeric:
    - Cari nilai mean juga standar deviasi variabelnya.
    - Cari data nilai probabilitistik. Caranya hitung data-data yang benar

## HASIL DAN PEMBAHASAN

### Algoritma C4.5

Pada tahapan ini dilakukan analisa perhitungan terhadap kasus permasalahan, dalam penelitian ini menggunakan Algoritma C4.5. Akan dijelaskan secara rinci detail Algoritma C4.5 menjadi analisa perhitungan, rule-rule yang dihasilkan dan gambaran pohon keputusan (decision tree). Dalam penelitian ini akan dilakukan eksperimen dengan menggunakan metode klasifikasi data mining Algoritma C4.5 dan Decision Tree terhadap data anak dengan gejala-gela yang dialaminya. Data akan diolah dengan menggunakan algoritma C4.5 dan menghasilkan model.

Dibuatnya data latih yang diambil dari rapidminer yakni:

**Tabel 2.** Data Latih

No	Produk	Kualitas	Label	Harga	Status
1	Palsu	Bagus	Ada	Rendah	Minat
2	Original	Tidak Bagus	Ada	Sedang	Minat
3	Original	Bagus	Ada	Tinggi	Tidak Minat
4	Original	Bagus	Ada	Sedang	Tidak Minat
5	Original	Tidak Bagus	Ada	Rendah	Minat
6	Original	Bagus	Ada	Sedang	Tidak Minat
7	Original	Tidak Bagus	Ada	Rendah	Minat
8	Original	Tidak Bagus	Ada	Sedang	Minat
9	Original	Bagus	Ada	Sedang	Minat
10	Original	Bagus	Ada	Tinggi	Tidak Minat
11	Palsu	Bagus	Ada	Rendah	Tidak Minat
12	Original	Bagus	Tidak Ada	Rendah	Minat
13	Palsu	Bagus	Ada	Rendah	Minat
14	Original	Bagus	Ada	Tinggi	Minat
15	Palsu	Bagus	Ada	Tinggi	Tidak Minat

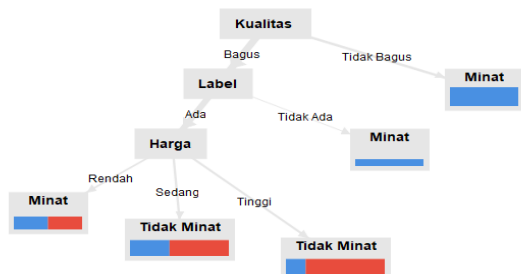
16	Original	Bagus	Ada	Sedang	Minat
17	Original	Tidak Bagus	Ada	Rendah	Minat
18	Original	Bagus	Tidak Ada	Rendah	Minat
19	Original	Bagus	Ada	Rendah	Tidak Minat
20	Original	Tidak Bagus	Ada	Rendah	Minat
21	Original	Bagus	Ada	Tinggi	Tidak Minat
22	Original	Bagus	Ada	Sedang	Tidak Minat

Dari tabel analisa perhitungan tersebut diatas, akan diperhitungkan nilai entropy dan gain dari masing-masing gejala.

**Tabel 3.** Model Perhitungan C4.5

		Jumlah (S)	Minat	Tidak Minat	Entropy	Gain
<b>Total</b>		22	13	9	0.976020648	
<b>Produk</b>						0.00541075
	Original	18	11	7	0.964078765	
	Palsu	4	2	2	1	
<b>Kualitas</b>						0.256966533
	Bagus	16	7	9	0.988699408	
	Tidak Bagus	6	6	0	0	
<b>Label</b>						0.073498417
	Ada	20	11	9	0.992774454	
	Tidak Ada	2	2	0	0	
<b>Harga</b>						0.170315268
	Rendah	10	8	2	0.721928095	
	Sedang	7	4	3	0.985228136	
	Tinggi	5	1	4	0.721928095	

Setelah itu, Perhitungan gain dan entropy dilakukan untuk seluruh karakter sudah selesai, sehingga didapatkan gain value paling tinggi. Seperti pada table 3 yang menampilkan hasil perhitungan semua variabel. Lalu dilanjutkan dengan menampilkan pohon keputusan akhir yang dibangun.



**Gambar 1.** Pohon Keputusan

Model yang telah dibentuk diuji tingkat akurasi dengan memasukan atau uji yang berasal dari data training dengan menggunakan cross validation pada aplikasi ripedminer untuk menguji tingkat akurasi. Selanjutnya dengan aplikasi rapidminer dihasilkan nilai akurasi, nilai class recall dan nilai class precision.

accuracy: 60.00%

	true Minat	true Tidak Minat	class precision
pred. Minat	1	0	100.00%
pred. Tidak Minat	2	2	50.00%
class recall	33.33%	100.00%	

**Gambar 2.** Nilai Akurasi Perhitungan C4.5

Dari gambar 2 diatas dijelaskan bahwa minat pelanggan menghasilkan class recall sebesar 33.33% dan class precision sebesar

100.00%, sedangkan tidak minat pelanggan menghasilkan class recall sebesar 100.00% dan class precision sebesar 50.00% dan nilai akurasi dari perhitungan algoritma C4.5 tersebut diatas adalah sebesar 60%.

### Algoritma Naïve Bayes

Pada tahapan ini dilakukan analisa perhitungan terhadap kasus permasalahan, dalam penelitian ini menggunakan Naïve Bayes. Akan dijelaskan secara rinci detail Naïve Bayes sebagai berikut:

**Tabel 4.** Model Perhitungan Naïve Bayes

P(P= I..	Minat	Tidak Minat
Original	85%	78%
Palsu	15%	22%
	<b>100%</b>	<b>100%</b>
P(K= I..	Minat	Tidak Minat
Bagus	54%	100%

Tidak Bagus	46%	0
	<b>100%</b>	<b>100%</b>
P(L= I..	Minat	Tidak Minat
Ada	85%	1
Tidak Ada	15%	0
	<b>100%</b>	<b>100%</b>
P(H= I..	Minat	Tidak Minat
Rendah	62%	22%
Sedang	31%	33%
Tinggi	8%	44%
	<b>100%</b>	<b>100%</b>

**Tabel 5.** Probabilitas Kelas Naïve Bayes

Probabilitas Kelas	
Status	Nilai
Minat	59%
Tidak Minat	41%
	100%

**Tabel 6.** Data Testing Naïve Bayes

Produk	Kualitas	Label	Harga	Status
Palsu	Bagus	Ada	Rendah	Minat
Original	Bagus	Ada	Tinggi	Minat
Palsu	Bagus	Ada	Tinggi	Tidak Minat
Original	Bagus	Ada	Sedang	Minat
Original	Tidak Bagus	Ada	Rendah	Minat

**Tabel 7.** Prediction Naïve Bayes

Class Prediction	Minat	Tidak Minat
Minat	2.55%	2.02%
TidakMinat	1.75%	14.14%
Minat	4.14%	4.04%
TidakMinat	7.01%	10.61%
Minat	12.02%	0.00%

**Tabel 8.** Confusion Table Naïve Bayes

Predicted	Class	
	Minat	Tidak Minat
Minat	2	1
Tidak Minat	2	0

Selanjutnya, model yang telah dibentuk diuji tingkat akurasi dengan memasukan atau uji yang berasal dari data training dengan menggunakan cross validation pada aplikasi

ripedminer untuk menguji tingkat akurasi. Dengan aplikasi rapidminer dihasilkan nilai akurasi, nilai class recall dan nilai class precision yang akan dijelaskan pada gambar dibawah ini:

accuracy: 60.00%

	true Minat	true Tidak Minat	class precision
pred. Minat	2	0	100.00%
pred. Tidak Minat	2	1	33.33%
class recall	50.00%	100.00%	

**Gambar 3.** Nilai Akurasi Perhitungan Naïve Bayes

Dari gambar 3 diatas dijelaskan bahwa minat pelanggan menghasilkan class recall sebesar 50.00% dan class precision sebesar 100.00%, sedangkan tidak minat pelanggan menghasilkan class recall sebesar 100.00% dan class precision sebesar 33.33% dan nilai akurasi

dari perhitungan algoritma Naive Bayes tersebut diatas adalah sebesar 60.00%.

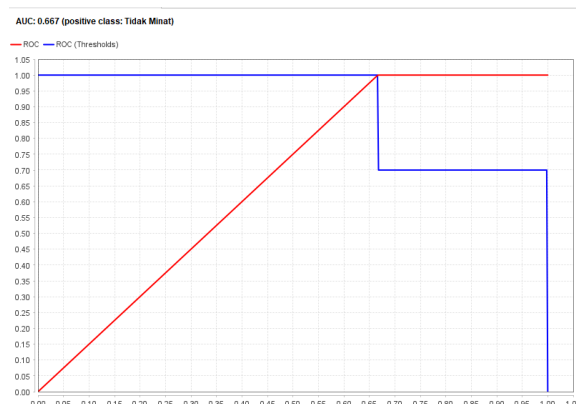
Setelah dua tabel Confusion matrix dibuat, hitung Accuracy, Precision, dan Recall. Selanjutnya menampilkan perbandingan nilai tersebut untuk metode C4.5 dan Naïve Bayes.

**Tabel 9.** Confusion Matrix

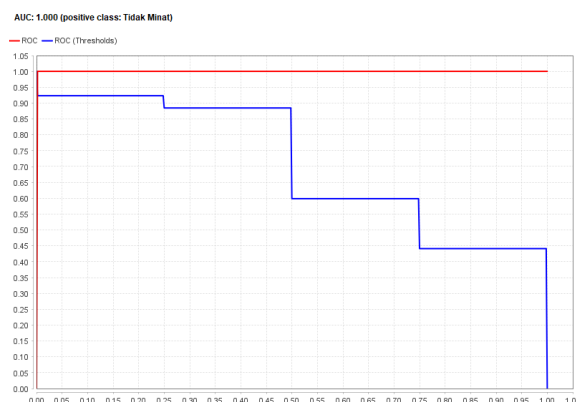
	C4.5	Naïve Bayes
<b>Accuracy</b>	60,00%	60,00%
<b>Precision</b>	50,00%	33,33%
<b>Recall</b>	100,00%	100,00%

Dari hasil table perbandingan diatas nilai akurasi algoritma C4.5 dan naïve bayes memiliki nilai akurasi yang sama dan sebanding, tetapi berbeda pada nilai presisi nya, algoritma C4.5 memiliki nilai presisi yang lebih tinggi dibandingkan nila presisi algoritma Naïve Bayes.

### Kurva ROC



**Gambar 4.** Kurva ROC dengan Algoritma C4.5



**Gambar 5.** Kurva ROC dengan Algoritma Naïve Bayes

Analisa Hasil Perbandingan Cross validation digunakan untuk menguji model hasil metode C4.5 dan Naïve Bayes.

**Tabel 10.** Cross Validation

	C4.5	Naïve Bayes
<b>Accuracy</b>	60,00%	60,00%
<b>AUC</b>	0,667	1,000

Dapat dilihat dari perbandingan nilai accuracy dan AUC, nilai accuracy C4.5 60,00%, namun nilai AUC dengan kriteria klasifikasinya yang buruk. Sedangkan, nilai accuracy Naïve Bayes memiliki nilai 60,00%, tapi nilai AUC termasuk ke dalam klasifikasi sangat baik.

### KESIMPULAN

Setelah beberapa langkah pengujian yang peneliti lakukan, maka peneliti dapat menarik kesimpulan dari hasil yang sudah didapatkan. Kesimpulan tersebut antara lain:

1. Berdasarkan seluruh dari hasil pengujian yang sudah dilakukan oleh peneliti didapatkan hasil pada algoritma C4.5 dan Naïve Bayes sama-sama memiliki nilai akurasi dan presisi tertinggi dengan data yang di uji adalah dataimbang. Algoritma C4.5 nilai akurasi sebesar 60,00% dan presisi sebesar 55,00% sedangkan algoritma Naïve Bayes nilai akurasi sebesar 60,00% dan presisi sebesar 33,33%.
2. Diketahui bahwa kedua algoritma memiliki hasil akurasi yang sama namun terdapat selisih nilai pada presisi. Namun dalam hal ini peneliti bertujuan untuk mengetahui nilai akurasi dan membandingkannya, sehingga algoritma C4.5 maupun algoritma Naïve Bayes sama-sama akurat dalam penentuan rekomendasi minat pelanggan. Disisi lain pengujian ini sama-sama terletak pada teknik uji dengan dataimbang.
3. Dari hasil semua pengujian peneliti lakukan kemudian membandingkan nilai akurasi dan didapatkan hasil yang bernilai sama sehingga dapat ditarik kesimpulan bahwa algoritma C4.5 dan Naïve Bayes sama-sama akurat, namun pada segi nilai presisi algoritma C4.5 lebih unggul dalam penentuan rekomendasi minat pelanggan.



## DAFTAR PUSTAKA

- Ainurrohmah, A. (2021). Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka. *PRISMA, Prosiding Seminar Nasional Matematika*, 493–499.
- Amalia, R. (2020). Penerapan data mining untuk memprediksi hasil kelulusan siswa menggunakan metode naïve bayes. *Jurnal Informatika Dan Sistem Informasi*, 6(1), 33–42.
- Aprilianur, G., & Hadisaputro, E. L. (2022). Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Analisa Penjualan Toko Myam Hijab Penajam. *JUPITER (Jurnal Penelitian Ilmu Dan Teknik Komputer)*, 14(1), 161–170.
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains Dan Pendidikan Informatika*, 2(2), 213–219.
- Peng, W., Huang, L., Jia, J., & Ingram, E. (2018). Enhancing the Naive Bayes Spam Filter Through Intelligent Text Modification Detection. *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, 849–854.  
<https://doi.org/10.1109/TrustCom/BigDataSE.2018.00122>
- Salim, Y., & Azis, H. (2018). Kinerja Metode C4.5 dalam Penyaluran Bantuan Dana Bencana. *Prosiding SAKTI (Seminar Ilmu Komputer Dan Teknologi Informasi)*, 84–87.
- Susanti, N. A., Walid, M., & Hoiriyah, H. (2022). Klasifikasi Data Tweet Ujaran Kebencian di Media Sosial Menggunakan Naive Bayes Classifier. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 538–543.  
<https://doi.org/10.36040/jati.v6i2.5174>