

# Data Mining Untuk Memprediksi Prestasi Peserta Didik di SMKN 1 Penawartama Tulang Bawang Menggunakan C4.5

Suroto<sup>\*1</sup>, Hendri Purnomo<sup>2</sup>, Randi Estian Pambudi<sup>3</sup>

<sup>1,2,3</sup>Jurusan Teknik Komputer, IIB Darmajaya

Jalan ZA. Pagar Alam No.93, Gedong Meneng, Kec. Rajabasa, Kota Bandar Lampung, Lampung 35141

e-mail: \*<sup>1</sup>[kangmashdy@gmail.com](mailto:kangmashdy@gmail.com), <sup>2</sup>[hendrialie@darmajaya.ac.id](mailto:hendrialie@darmajaya.ac.id),  
<sup>3</sup>[randiestian@darmajaya.ac.id](mailto:randiestian@darmajaya.ac.id)

## Abstrak

*SMK Negeri 1 Penawartama Tulang Bawang adalah sebuah sekolah yang siswanya berasal dari berbagai desa di kabupaten tersebut. Mayoritas siswa berasal dari keluarga dengan kondisi ekonomi yang kurang mampu dan tingkat pendidikan yang rendah. Kondisi ini sering memengaruhi motivasi belajar siswa. Penelitian ini bertujuan untuk memprediksi prestasi akademik siswa berdasarkan status sosial ekonomi orang tua, tingkat motivasi, tingkat kedisiplinan, dan prestasi akademik sebelumnya dengan menggunakan metode data mining dan algoritma C4.5. Penelitian ini menggunakan pendekatan kuantitatif dengan melibatkan 606 siswa kelas X di SMK Negeri 1 Penawartama Tulang Bawang. Data dikumpulkan melalui dokumentasi dan angket. Hasil penelitian menunjukkan bahwa analisis prediksi dengan pohon keputusan memiliki tingkat akurasi 98,02%, presisi 94,44%, dan recall 77,27%.*

**Kata kunci:** data mining, algoritma C4.5, akurasi. Presisi dan recall

## Abstract

*SMK Negeri 1 Penawartama Tulang Bawang is a school whose students come from various villages in the district. Most of the students come from families with limited economic conditions and low levels of education. These factors often affect the students' learning motivation. The aim of this research is to predict students' academic performance based on parents' socioeconomic status, motivation level, discipline level, and previous academic achievements using data mining methods with the C4.5 algorithm. This research employs a quantitative approach involving 606 tenth-grade students at SMK Negeri 1 Penawartama Tulang Bawang. Data collection methods used include documentation and questionnaires. The research results show that the prediction analysis using decision trees has an accuracy rate of 98.02%, precision of 94.44%, and recall of 77.27%.*

**Keywords:** data mining, C4.5 algorithm, accuracy, precision, and recall

## 1. PENDAHULUAN

**S**ekolah Menengah Kejuruan (SMK) adalah salah satu bentuk institusi pendidikan formal yang menyediakan pendidikan vokasional di tingkat menengah. Pendidikan vokasional ini bertujuan untuk memenuhi kebutuhan masyarakat akan tenaga kerja yang terampil[1][2]. Program pendidikan dan pelatihan di SMK bertujuan untuk mempersiapkan siswa agar memiliki pemahaman, keterampilan, pengembangan diri, perilaku, sikap, kebiasaan kerja, dan apresiasi

terhadap pekerjaan yang dibutuhkan oleh dunia usaha/industri. Proses ini diawasi oleh pemerintah dan masyarakat atau berdasarkan kontrak dengan lembaga tertentu serta berbasis pada produktivitas [3].

SMK Negeri 1 Penawartama Tulang Bawang merupakan salah satu sekolah paling diminati di Kota Tulang Bawang. Salah satu program keahlian yang ada di SMK Negeri 1 Penawartama Tulang Bawang adalah program keahlian multimedia. Program keahlian multimedia memiliki fokus keahlian dalam bidang desain grafis, animasi, dan fotografi yang membekali siswa untuk bekerja di dalam bidang animasi, grafis percetakan, pembuatan multimedia interaktif secara inovatif dan profesional. Hasil belajar merupakan perubahan yang diterima oleh seorang manusia ketika belajar. Hasil belajar dapat berupa perubahan dalam kemampuan kognitif, afektif, dan psikomotorik tergantung dari tujuan belajarnya [4]. Hasil belajar siswa berbeda tergantung dalam kondisi masing-masing siswa. Berbagai faktor seperti kondisi keluarga, sekolah, minat, motivasi, dan prestasi masa lalu menjadi penentu pencapaian hasil belajar.

Masalah yang dihadapi oleh SMK Negeri 1 Penawartama Tulang Bawang adalah bahwa siswanya berasal dari berbagai desa di kabupaten tersebut. Mayoritas dari siswa berasal dari keluarga dengan latar belakang ekonomi dan pendidikan yang rendah. Selain itu, motivasi belajar siswa cenderung rendah karena pengaruh lingkungan keluarga dan masyarakat sekitar. Sebagian besar siswa di sekolah ini tidak memiliki prestasi yang memuaskan, baik dalam bidang akademik maupun non-akademik. Penambangan data merupakan istilah yang digunakan untuk menggambarkan proses menemukan pengetahuan dalam kumpulan data. Proses ini melibatkan penggunaan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi serta pengetahuan yang berguna dari data dalam jumlah besar. Saat ini, penambangan data banyak dimanfaatkan di bidang pendidikan untuk mendukung pengambilan keputusan berdasarkan analisis data guna meningkatkan kualitas pembelajaran [5]. Algoritma C4.5 merupakan salah satu teknik dalam bidang data *mining* yang berfokus pada pembentukan pohon keputusan. Pohon keputusan ini digunakan untuk melakukan klasifikasi dan prediksi dengan efektif. Metode ini mengubah data yang kompleks menjadi struktur pohon yang merepresentasikan aturan-aturan spesifik. Lebih lanjut, pohon keputusan juga berguna untuk mengeksplorasi data dan mengungkapkan keterkaitan tersembunyi antara berbagai variabel input dan target [6].

Sejumlah studi telah dilaksanakan dengan menerapkan teknik penambangan data untuk meramalkan pencapaian akademis siswa. Penggunaan teknik ini bertujuan untuk memperkirakan prestasi belajar berdasarkan variabel-variabel seperti kualitas pengajaran, motivasi, kedisiplinan, kondisi ekonomi, dan hasil akademis. Variabel yang dianalisis termasuk IPK, faktor sosial, tingkat disiplin, motivasi, dan nilai rapor, dengan tingkat akurasi mencapai 65%.[7]. Penerapan Teknik Data *Mining* dengan Metode *Naïve Bayes* untuk meramalkan prestasi akademis siswa di Sekolah Menengah Pertama (SMP), dengan mempertimbangkan faktor-faktor seperti jumlah anggota keluarga, status dan pendidikan orang tua, pekerjaan orang tua, kehadiran pengasuh, jarak rumah ke sekolah, waktu belajar, dukungan keluarga, serta partisipasi dalam kegiatan ekstrakurikuler, mencapai tingkat ketepatan prediksi sebesar 56,7%.[8].

Terdasarkan pada penjelasan sebelumnya, algoritma *data mining* C4.5 memiliki aplikasi potensial dalam meramalkan pencapaian akademik siswa di kelas X Multimedia di SMK Negeri 1 Penawartama Tulang Bawang. Implementasi ini dapat membantu pengajar dalam mengidentifikasi siswa dengan pencapaian yang tidak optimal. Studi lebih lanjut diperlukan untuk memahami penerapan algoritma C4.5 dalam konteks evaluasi prestasi siswa kelas X Multimedia.

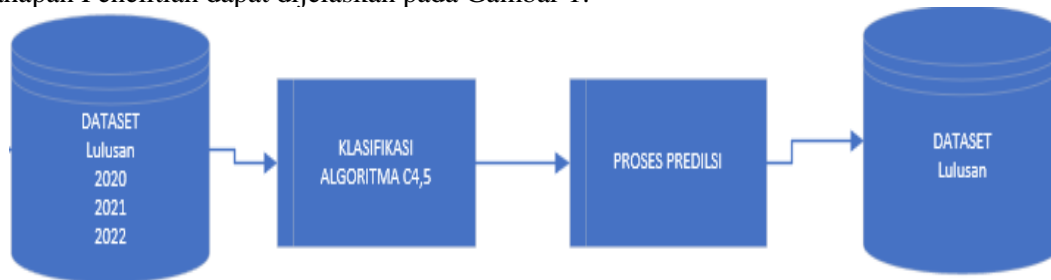
## 2. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dan dilaksanakan di SMK Negeri 1 Penawartama, Tulang Bawang. Prosedur penelitian mengikuti tahapan-tahapan dalam Knowledge Data Discovery (KDD). Langkah-langkahnya adalah sebagai berikut:

1. Pre-processing/Cleaning: Sebelum dilakukan data mining, data dibersihkan dengan menghapus duplikasi, memeriksa inkonsistensi, dan memperbaiki kesalahan seperti kesalahan tipografi;
2. Transformation: Data yang telah dipilih diubah melalui pengkodean agar siap digunakan dalam proses data mining;
3. Analisis data: Data dianalisis dengan cara diperiksa, dibersihkan, diubah, dan dimodelkan untuk menemukan informasi yang berguna, menarik kesimpulan, dan mendukung pengambilan keputusan;
4. Pattern Evaluation: Pada tahap ini, pola-pola yang menarik dievaluasi untuk memastikan mereka benar-benar mewakili pengetahuan dari data yang ada.

### 2.1 Tahapan Pengolahan Data

Tahapan Penelitian dapat dijelaskan pada Gambar 1.



Gambar 1. Tahapan Pengolahan Data

1. Dataset lulusan adalah kumpulan data yang terorganisir, biasanya dalam bentuk tabel atau file elektronik, yang digunakan untuk analisis atau pembelajaran mesin. Dataset biasanya terdiri dari baris dan kolom yang mewakili *instance* atau observasi data serta variabel atau fitur yang menggambarkan karakteristik dari setiap *instance* tersebut. Dataset digunakan dalam berbagai bidang, termasuk ilmu data, statistik, pembelajaran mesin, dan penelitian lainnya untuk mengeksplorasi pola, membuat prediksi, atau menguji hipotesis.
2. Algoritma klasifikasi yang digunakan dalam data *mining* dan *machine learning* untuk membangun *decision tree* (pohon keputusan)
3. Proses prediksi ini penting dalam konteks penelitian untuk mengukur kinerja model dan relevansi informasi yang diperoleh dari data yang tersedia

### 2.2 Teknik Analisis Data

Penelitian ini menggunakan teknik *decision tree*, *CHAID*, dan regresi ganda untuk memprediksi prestasi belajar siswa SMK 1 Penawartama Tulang Bawang, dengan variabel-variabel yang meliputi status ekonomi orang tua, motivasi, kedisiplinan, dan prestasi masa lalu. *Decision Tree* akan menampilkan faktor-faktor yang kemungkinan mempengaruhi prestasi belajar siswa, serta memberikan prediksi hasil akhir jika faktor-faktor tersebut terpenuhi. Diagram pohon akan menggambarkan visualisasi data dalam bentuk aturan keputusan. Proses analisis *Decision Tree* meliputi *transformasi* data menjadi model pohon, konversi model pohon menjadi aturan, dan penyederhanaan aturan melalui *pruning*. Penelitian ini melibatkan 416 siswa sebagai

populasi untuk pembuatan model prediksi *Decision Tree*, yang akan dievaluasi untuk tingkat akurasi prediksinya.[9].

### 2.3 CHAID

Metode ini bertujuan untuk mengelompokkan data secara berurutan dengan membagi data menjadi beberapa subgrup berdasarkan pembagian biner. Pada setiap tahapnya, pengelompokan grup dilakukan dengan membagi grup menjadi dua bagian menggunakan salah satu variabel prediktor. Setiap bagian dari kategori variabel tersebut mendefinisikan satu subgrup, sementara kategori yang tersisa membentuk subgrup lainnya. Dalam *CHAID*, prediktor dapat berupa tipe monotonik atau bebas[10]. Proses analisis data dengan *CHAID* dimulai dengan menguji chi-square untuk setiap variabel independen, menentukan variabel independen yang paling signifikan, membagi data berdasarkan kategori variabel independen tersebut dengan signifikansi tertinggi, dan mengulangi langkah ini untuk semua subgrup sampai semua pembagian yang signifikan secara statistik teridentifikasi.

### 2.4 Regresi

*Linier* adalah teknik statistik yang digunakan untuk membangun model yang menghubungkan variabel terikat (dependen) dengan satu atau lebih variabel independen. Ketika hanya ada satu variabel independen, disebut sebagai regresi linier sederhana. Namun, jika terdapat lebih dari satu variabel independen[11]. Analisis regresi memiliki minimal tiga manfaat utama. Pertama, digunakan untuk menjelaskan fenomena data atau kasus yang sedang diteliti. Kedua, berguna untuk mengontrol variabel dalam sebuah kasus atau situasi tertentu dengan menggunakan model regresi yang telah dikembangkan. Ketiga, dapat digunakan untuk melakukan prediksi terhadap variabel dependen. Namun, penting untuk diingat bahwa prediksi dalam konteks regresi hanya dapat dilakukan dalam rentang data yang digunakan untuk membentuk model regresi tersebut [12].

### 2.5 Algoritma C4.5

Secara umum Algoritma C4.5 untuk membangun pohon keputusan seperti pada penjelasan ini.

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

Keterangan: Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan rumus seperti berikut[13]:

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

S: himpunan kasus

A: atribut

N: jumlah partisi atribut

A |S<sub>i</sub> |: jumlah kasus pada partisi ke-i

|S|: jumlah kasus dalam S

Sebelum mendapatkan nilai *Gain* adalah dengan mencari nilai *Entropy*. *Entropy* digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar dari *Entropy* adalah sebagai berikut:

$$Entropy S = -\sum p_i * \log_2 p_i \quad (2)$$

Keterangan:

S: himpunan kasus

A: fitur

n: jumlah partisi

S<sub>Pi</sub>: proporsi dari S<sub>i</sub> terhadap S

Contoh perhitungan

Dari kutipan di atas Perhitungan C4.5 dimulai dengan mencari nilai *entropy* dari semua data. Nilai *entropy* digunakan sebagai dasar perhitungan gain tiap atribut, nilai *gain* tertinggi menjadi akar dari pohon keputusan perhitungan dilakukan hingga semua atribut terdefinisi [14].

## 2.6 Split Info

*Split Info* merupakan rumus yang menyatakan informasi potensial atau *entropy*. dapat dilihat dalam rumus (3). Dan keterangannya:

$$\text{Split Info}(S, A) = -\sum_{i=1}^n S_i \log_2 S_i \quad (3)$$

S : Himpunan kasus

A : Atribut

S<sub>i</sub> : Jumlah kasus pada partisi ke- i

## 2.7 Gain Ratio

*Gain Ratio* adalah modifikasi dari information gain yang digunakan untuk mengurangi bias atribut yang memiliki banyak cabang. *Gain ratio* memiliki sifat:

Bernilai besar jika data menyebar rata

Bernilai kecil jika semua data masuk ke dalam satu cabang *Gain Ratio*

$$\text{Gain Ratio}(S, A) = \text{Gain}(S, A) \cdot \text{SplitInfo}(S, A) \quad (4)$$

dan Keterangannya sebagai berikut:

S: Himpunan kasus

A: Atribut Gain

(S, A): Information gain pada atribut A

SplitInfo (S, A): SplitInfo pada atribut A

$$\text{Akurasi} = \frac{(TN+TP)}{(TN+FN+FP+TP)} \quad (5)$$

*Precision* digunakan untuk mengukur seberapa besar proporsi dari kelas data positif yang berhasil diprediksi dengan benar dari keseluruhan hasil prediksi kelas positif.

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

*Recall* digunakan untuk menunjukkan presentase kelas data positif yang berhasil diprediksi benar dari keseluruhan data kelas positif.

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (7)$$

## 2.8 Pengujian Algoritma Decision Tree C4.5

Pengujian data yang digunakan dalam algoritma C4.5 adalah dataset prediksi bantuan beasiswa yang didapatkan dari data 3 tahun sebelumnya [15].

### 3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, data yang digunakan adalah dataset yang mencakup data private catatan siswa dari SMKN 1 Penawartama untuk periode 2018-2021. 600 Data tersebut mencakup 7 variabel seperti NISN, Nama, Nilai, Status Ekonomi, Kedisiplinan, Presensi, dan Prestasi.

#### 3.1. Transformasi Status Ekonomi

Transformasi ini digunakan untuk menentukan range Status Ekonomi, dengan kategori seperti pada tabel 1.

Tabel 1 Transformasi Status Ekonomi

Status Ekonomi	Range
Kurang	< 1.500.000
Cukup	1.500.000-2.500.000
Mampu	>2.500.000

Keterangan Tabel:

a. Kurang:

Range Penghasilan: Kurang dari Rp 1.500.000

Deskripsi: Kategori ini mencakup individu atau keluarga yang memiliki penghasilan bulanan di bawah Rp 1.500.000. Siswa mungkin menghadapi kesulitan dalam memenuhi kebutuhan dasar sehari-hari seperti makanan, tempat tinggal, dan pendidikan.

b. Cukup:

Range Penghasilan: Rp 1.500.000 - Rp 2.500.000

Deskripsi: Kategori ini mencakup individu atau keluarga yang memiliki penghasilan bulanan antara Rp 1.500.000 hingga Rp 2.500.000. Siswa biasanya dapat memenuhi kebutuhan dasar sehari-hari dan mungkin memiliki sedikit sisa untuk tabungan atau keperluan lain.

c. Mampu:

Range Penghasilan: Lebih dari Rp 2.500.000

Deskripsi: Kategori ini mencakup individu atau keluarga yang memiliki penghasilan bulanan lebih dari Rp 2.500.000. Siswa cenderung memiliki kehidupan yang lebih stabil secara finansial, mampu memenuhi kebutuhan dasar dan memiliki lebih banyak fleksibilitas untuk tabungan, investasi, atau pengeluaran tambahan lainnya

#### 3.2. Transformasi Kedisiplinan

Transformasi ini digunakan untuk menentukan range Kedisiplinan, dengan kategori seperti pada tabel 2.

Tabel 2 Transformasi Kedisiplinan

Kedisiplinan	Range
Kurang Disiplin	Tidak mematuhi tata tertib >70%
Cukup Disiplin	Kadang-kadang mematuhi tata tertib 70%
Disiplin	Selalu mematuhi tata tertib 70%

- Kurang Disiplin: Individu yang termasuk dalam kategori ini tidak mematuhi tata tertib lebih dari 70% waktu.
- Cukup Disiplin: Individu yang termasuk dalam kategori ini kadang-kadang mematuhi tata tertib sekitar 70% waktu.
- Disiplin: Individu yang termasuk dalam kategori ini selalu mematuhi tata tertib 70% waktu.

### 3.3. Transformasi Presensi

Transformasi ini proses memperkirakan atau memproyeksikan kinerja akademik atau pencapaian siswa di masa depan berdasarkan data dan informasi yang ada. Prediksi ini bisa dilakukan dengan berbagai metode, termasuk analisis statistik, machine learning, atau metode prediktif lainnya. Berikut adalah beberapa faktor yang sering digunakan dalam memprediksi prestasi peserta didik: digunakan untuk menentukan range Presensi, dengan kategoriseperti pada tabel 3.

Tabel 3 Transformasi Presensi

Presensi	Range
Tidak Rajin	Kehadiran < 70%
Rajin	Kehadiran > 70%

- Tidak Rajin: Jika kehadiran kurang dari 70%.
- Rajin: Jika kehadiran lebih dari 70%.

Untuk menemukan *entropi* dan gain guna menetapkan simpul *root* awal, berdasarkan proses sebelumnya, ditemukan bahwa menggunakan alat *RapidMiner*, simpul root awalnya adalah usia. Di sini, kami akan mengukur nilai entropi dan gain secara manual. Langkah ini bertujuan untuk menguji keefektifan alat *RapidMiner*.

Hasil akurasi dari *Decision Tree C4.5* dan *validasi Split*, dengan data sampel 70% dan data uji 98.64%, menunjukkan akurasi sebesar 98,02%.

Menghitung akurasi:

$$\text{Akurasi} = \frac{(561+34)}{(561+34+10+2)} = \frac{595}{607} = 98,02\%$$

Berikut adalah hasil *precision* yaitu mendapatkan 94,44%

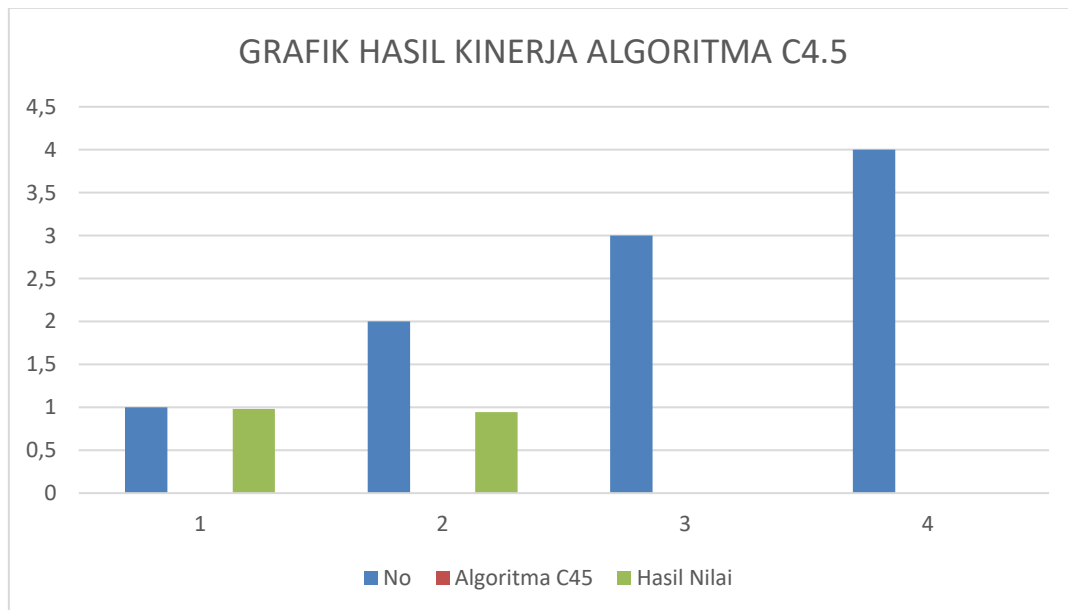
Menghitung *Precision*,

$$:p(1) = \frac{561}{(561+10)} = \frac{561}{571} = 98,25 \quad :p(0) = \frac{34}{(34+2)} = \frac{34}{36} = 94,44\%$$

Berikut adalah hasil *Recall* yaitu mendapatkan 77,27 %

Menghitung *Recall*,

$$R(1) = \frac{561}{(561+2)} = \frac{561}{569} = 99,54\% \quad R(0) = \frac{34}{(34+10)} = \frac{34}{44} = 77,27\%$$



Gambar 2 Grafik hasil Kinerja Algoritma C4.5

Gambar ini menunjukkan tiga metrik utama yang digunakan untuk mengevaluasi kinerja model klasifikasi: Akurasi, *Recall*, dan *Precision*.

1. **Akurasi:** Metrik ini mengukur seberapa sering model membuat prediksi yang benar secara keseluruhan. Akurasi dihitung sebagai rasio prediksi benar (baik positif maupun negatif) terhadap total jumlah prediksi. Nilai akurasi yang tinggi menunjukkan bahwa model memiliki performa yang baik dalam mengidentifikasi kelas yang benar secara keseluruhan.
2. ***Recall (Sensitivity):*** Metrik ini mengukur seberapa baik model mengenali sampel positif. *Recall* dihitung sebagai *rasio true positives (TP)* terhadap jumlah sebenarnya dari sampel positif ( $TP + FN$ ). *Recall* yang tinggi berarti model jarang melewatkan sampel yang seharusnya diklasifikasikan sebagai positif.
3. ***Precision:*** Metrik ini mengukur seberapa baik model dalam memastikan bahwa sampel yang diprediksi sebagai positif memang benar-benar positif. *Precision* dihitung sebagai *rasio true positives (TP)* terhadap jumlah total prediksi positif ( $TP + FP$ ). *Precision* yang tinggi menunjukkan bahwa model jarang membuat kesalahan dengan mengklasifikasikan sampel negatif sebagai positif.

Dari gambar ini, kita bisa melihat bahwa model memiliki akurasi yang tinggi, yang berarti sebagian besar prediksi model adalah benar. *Recall* yang cukup tinggi menunjukkan bahwa model cukup baik dalam mendeteksi sampel positif, meskipun masih ada beberapa sampel positif yang terlewat. *Precision* yang tinggi menunjukkan bahwa sebagian besar prediksi positif model adalah benar, dengan sedikit kesalahan prediksi positif.



#### 4. KESIMPULAN

Berdasarkan hasil yang diberikan, *algoritma C4.5* menunjukkan kinerja yang sangat baik dalam mengklasifikasikan data. Dengan tingkat akurasi mencapai 98,02%, presisi sebesar 94,44%, dan *recall* atau *matriks* kebingungan sebesar 77,72%, algoritma ini menunjukkan kemampuan yang solid dalam mengidentifikasi dan mengklasifikasikan data secara tepat. Secara keseluruhan, hasil ini menunjukkan bahwa *algoritma C4.5* cocok digunakan dalam kasus pengklasifikasian yang rumit dan memerlukan presisi yang tinggi.

#### 5. SARAN

Penelitian selanjutnya, disarankan untuk menjelajahi beberapa arah yang dapat mengembangkan pengetahuan dalam bidang *machine learning*. Salah satu pendekatan dapat fokus pada pengembangan model *hybrid* yang menggabungkan keunggulan berbagai teknik seperti *ensemble learning* atau *deep learning* dengan metode tradisional seperti regresi atau klasifikasi. Eksplorasi dalam domain aplikasi khusus seperti biomedis atau keuangan juga dapat memberikan wawasan tentang bagaimana *machine learning* dapat digunakan untuk memecahkan tantangan kritis dalam berbagai industri.

#### DAFTAR PUSTAKA

- [1] P. Sudira, "Technology and Vocational Education," *Appl. Sci. Educ. Bandung Pedagog. Press.*, pp. 1285–1300, 2007.
- [2] W. AKBAR, "IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA C4. 5 UNTUK MEMPREDIKSI PENJUALAN ALAT–ALAT TELEKOMUNIKASI DAN IT (Information ...," pp. 597–606, 2022, [Online]. Available: <https://repository.binadarma.ac.id/id/eprint/7334%0Ahttp://repository.binadarma.ac.id/7334/2/67312-6.pdf>
- [3] H. Ponto, N. M. Sangi, and C. T. M. Manopo, "Developing students' science process skills in the learning of electricity," *Int. J. Innov. Creat. Chang.*, vol. 5, no. 6, pp. 1–18, 2019.
- [4] A. Purwanto, Y. Yusralaini, and S. Susilatri, "Pengaruh Likuiditas, Leverage, Manajemen Laba, Dan Kopersasi Rugi Fiskal Terhadap Agresivitas Pajak Perusahaan Pada Perusahaan Pertanian Dan Pertambangan Yang Terdaftar Di Bursa Efek Indonesia Periode 2011-2013," *J. Online Mhs. Fak. Ekon. Univ. Riau*, vol. 3, no. 1, pp. 580–594, 2016.
- [5] S. Aydogdu, "Educational data mining studies in Turkey: A systematic review," *Turkish Online J. Distance Educ.*, vol. 21, no. 3, pp. 170–185, 2020, doi: 10.17718/TOJDE.762046.
- [6] G. C. Sutradana and M. D. R. Wahyudi, "Penerapan Data Mining Untuk Analisis Pengaruh Lama Studi Mahasiswa Teknik Informatika Uin Sunan Kalijaga Yogyakarta Menggunakan Metode Apriori," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 3, p. 153, 2017, doi: 10.14421/jiska.2017.13-07.
- [7] E. Sabna and M. Muhandi, "Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi, dan Hasil Belajar," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 2, no. 2, p. 41, 2016, doi: 10.24014/coreit.v2i2.2392.
- [8] F. Rahman, D. Muhammad, and I. Firdaus, "Penerapan Data Mining Metode Naïve Bayes Untuk Prediksi Hasil Belajar Siswa Sekolah Menengah Pertama (Smp)," *Al Ulum Sains dan Teknol.*, vol. 1, no. 2, pp. 76–78, 2016.
- [9] I. Budiman and R. Ramadina, "Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi," *Ijccs*, vol. x, No.x, no. 1, pp. 1–5, 2015.

- [10] R. S. Rahayu, M. A. Mukid, and T. Wuryandari, “Identifikasi faktor-faktor yang mempengaruhi terjadinya preeklampsia dengan metode CHAID,” *J. Gaussian*, vol. 4, no. 2000, pp. 383–392, 2015.
- [11] S. C. Wohon, D. Hatidja, and N. Nainggolan, “Penentuan Model Regresi Terbaik Dengan Menggunakan Metode Stepwise ( Studi Kasus : Impor Beras Di Sulawesi Utara ) Determining the Best Regression Model Using Stepwise Method ( Case Study : Rice Imports in North Sulawesi ),” *J. Ilm. Sains*, vol. 17, no. 2, p. 81, 2017.
- [12] Syilfi, D. Ispriyanti, and D. Safitri, “Analisis Regresi Linier Piecewise Dua Segmen,” *J. Gaussian*, vol. 1, no. 1, pp. 219–228, 2012.
- [13] B. Novianti, T. Rismawan, and S. Bahri, “Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak),” *J. Coding, Sist. Komput. Untan*, vol. 04, no. 3, pp. 75–84, 2016.
- [14] M. Mustika, E. P. A. Sugara, and M. Pratiwi, “Pengembangan Media Pembelajaran Interaktif dengan Menggunakan Metode Multimedia Development Life Cycle,” *J. Online Inform.*, vol. 2, no. 2, p. 121, 2018, doi: 10.15575/join.v2i2.139.
- [15] T. Wahyudi *et al.*, “Penerapan Data Mining Pada Transaksi Penjualan Pakaian Dengan Menggunakan Algoritma Apriori,” *Jupiter*, vol. 14, no. 2, pp. 473–482, 2022.