



Penerapan Algoritma C4.5 untuk Klasifikasi Tingkat Kedisiplinan Siswa Sekolah Menengah

Seli Puri Andani^{1,*}, Rosyana Fitria Purnomo², Yodhi Yuniarthe²

¹ Fakultas Teknik, Prodi Informatika, Universitas Muhammadiyah Lampung, Bandar Lampung, Indonesia

² Fakultas Komputer, Prodi Sistem Informasi, Universitas Mitra Indonesia, Bandar Lampung, Indonesia

² Fakultas Komputer, Prodi Informatika, Universitas Mitra Indonesia, Bandar Lampung, Indonesia

Email: selipuria@gmail.com¹, rosyana.purnomo@umitra.ac.id², yodhi@umitra.ac.id³

Email Penulis Korespondensi: selipuria@gmail.com

Abstrak—Penelitian ini bertujuan untuk menguji kinerja algoritma *Decision Tree* berbasis kriteria *entropy* (C4.5) dalam mengklasifikasikan kelayakan mahasiswa berdasarkan data akademik maupun non-akademik. Dataset yang digunakan terdiri dari 200 entri dengan sembilan atribut, meliputi persentase kehadiran, jumlah keterlambatan, pelanggaran disiplin, nilai rata-rata, partisipasi, jam belajar, serta kegiatan ekstrakurikuler. Proses pengolahan data dilakukan melalui tahap pembersihan, transformasi, pemilihan atribut, pembagian data latih dan uji, hingga evaluasi model menggunakan confusion matrix. Hasil pengujian menunjukkan bahwa model yang dibangun mampu menghasilkan akurasi sebesar 87,5%, presisi rata-rata 85,6%, recall rata-rata 84,2%, dan F1-Score sebesar 84,8%. Temuan ini menegaskan bahwa algoritma C4.5 dapat digunakan secara efektif dalam mendukung klasifikasi performa mahasiswa dengan tingkat keandalan yang cukup tinggi.

Kata Kunci: Data Mining; Klasifikasi; Decision Tree; C4.5; Evaluasi Model; Performa Mahasiswa

Abstract—This study aims to evaluate the performance of the *Decision Tree* algorithm based on the *entropy* criterion (C4.5) in classifying student eligibility by considering both academic and non-academic data. The dataset consists of 200 entries with nine attributes, including attendance percentage, number of lateness incidents, disciplinary violations, average academic scores, participation, study hours, and extracurricular activities. Data processing was carried out through several stages, namely cleaning, transformation, feature selection, training and testing data splitting, and model evaluation using a confusion matrix. The experimental results show that the proposed model achieved an accuracy of 87.5%, an average precision of 85.6%, an average recall of 84.2%, and an F1-Score of 84.8%. These findings confirm that the C4.5 algorithm can be effectively applied to support student performance classification with a fairly high level of reliability.

Keywords: Data Mining; Classification; Decision Tree; C4.5; Model Evaluation; Student Performance

1. PENDAHULUAN

Kedisiplinan merupakan salah satu faktor penting yang menentukan keberhasilan siswa dalam mencapai prestasi akademik maupun non-akademik. Disiplin tidak hanya berkaitan dengan kepatuhan terhadap tata tertib sekolah, tetapi juga mencakup aspek tanggung jawab, manajemen waktu, serta sikap menghargai aturan. Siswa yang memiliki tingkat kedisiplinan tinggi cenderung menunjukkan prestasi belajar yang lebih baik, sedangkan rendahnya disiplin dapat memicu berbagai permasalahan, seperti keterlambatan, ketidakhadiran, hingga penurunan hasil belajar. Oleh karena itu, analisis terhadap tingkat kedisiplinan siswa menjadi penting untuk mendukung pihak sekolah dalam pengambilan keputusan yang tepat [1][2].

Seiring perkembangan teknologi informasi, data siswa yang terkumpul di sekolah semakin banyak dan beragam, mulai dari data absensi, nilai akademik, catatan pelanggaran, hingga keterlibatan dalam kegiatan ekstrakurikuler. Data yang besar ini seringkali belum dimanfaatkan secara optimal untuk memberikan informasi yang bermakna. Padahal, melalui pengolahan data yang tepat, sekolah dapat memperoleh wawasan yang berguna untuk mendeteksi pola kedisiplinan siswa. Di sinilah peran data mining menjadi relevan, karena mampu mengekstraksi pengetahuan tersembunyi dari data yang tersedia [3][4].

Data mining merupakan salah satu bidang dalam kecerdasan buatan yang berfokus pada proses menemukan pola dan informasi penting dari data berukuran besar. Teknik data mining telah banyak digunakan dalam dunia pendidikan melalui pendekatan *Educational Data Mining* (EDM). Dengan EDM, sekolah dapat melakukan prediksi, klasifikasi, maupun clustering terhadap data siswa untuk menghasilkan rekomendasi kebijakan yang lebih terarah. Salah satu algoritma yang sering digunakan dalam klasifikasi adalah algoritma C4.5, karena memiliki keunggulan dalam menghasilkan pohon keputusan yang mudah dipahami serta interpretatif [5][6].

Algoritma C4.5 bekerja dengan cara membangun pohon keputusan berdasarkan atribut-atribut yang paling berpengaruh terhadap kelas yang diprediksi. Dalam konteks kedisiplinan siswa, atribut seperti jumlah keterlambatan, tingkat kehadiran, catatan pelanggaran tata tertib, maupun partisipasi dalam kegiatan sekolah dapat dijadikan sebagai parameter untuk mengklasifikasikan tingkat disiplin siswa. Hasil klasifikasi ini dapat digunakan guru, wali kelas, maupun pihak manajemen sekolah untuk melakukan intervensi lebih dini kepada siswa yang berpotensi mengalami penurunan disiplin [7][8].

Beberapa penelitian sebelumnya telah mengkaji penerapan algoritma C4.5 pada bidang pendidikan. Penelitian oleh [9] menunjukkan bahwa metode C4.5 efektif digunakan dalam menentukan siswa teladan berdasarkan aturan kedisiplinan di sekolah. Hasilnya membuktikan bahwa algoritma ini mampu menghasilkan aturan klasifikasi yang mudah dipahami



oleh pihak sekolah. Temuan serupa juga dijelaskan dalam studi [10] yang memanfaatkan atribut kedisiplinan sebagai salah satu faktor penting dalam memprediksi prestasi siswa SMK, dengan tingkat akurasi yang cukup tinggi.

Selain itu, [11] menerapkan algoritma C4.5 untuk mengklasifikasikan predikat capaian siswa di madrasah dan berhasil membangun model yang dapat membantu pihak sekolah dalam proses evaluasi. Penelitian lain oleh [12] juga menekankan bahwa variabel kedisiplinan memiliki kontribusi signifikan dalam memprediksi prestasi akademik siswa, sehingga relevan untuk digunakan sebagai variabel utama dalam penelitian klasifikasi.

Temuan-temuan tersebut memperlihatkan bahwa algoritma C4.5 tidak hanya akurat dalam mengolah data siswa, tetapi juga memberikan aturan klasifikasi yang transparan dan dapat dijadikan dasar pengambilan keputusan oleh pihak sekolah. Oleh karena itu, penelitian ini berfokus pada penerapan algoritma C4.5 untuk mengklasifikasikan tingkat kedisiplinan siswa sekolah menengah, sehingga dapat mendukung upaya sekolah dalam meningkatkan kualitas kedisiplinan peserta didik.

Beberapa penelitian sebelumnya telah membuktikan bahwa algoritma C4.5 efektif dalam mengklasifikasikan data pendidikan, baik untuk prediksi prestasi akademik, identifikasi siswa berisiko drop out, maupun analisis perilaku belajar. Namun, penerapan algoritma ini secara khusus untuk klasifikasi kedisiplinan siswa sekolah menengah masih relatif jarang dilakukan. Hal ini membuka peluang penelitian untuk mengkaji lebih dalam bagaimana algoritma C4.5 dapat membantu sekolah dalam mengelompokkan tingkat kedisiplinan siswa secara lebih akurat dan sistematis [13][14].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam klasifikasi tingkat kedisiplinan siswa sekolah menengah. Dengan adanya hasil klasifikasi, diharapkan sekolah dapat memiliki acuan dalam menyusun strategi pembinaan dan pengawasan yang lebih tepat sasaran. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan *Educational Data Mining* di Indonesia, khususnya dalam bidang analisis perilaku siswa yang berhubungan dengan kedisiplinan.

2. METODOLOGI PENELITIAN

Penelitian ini termasuk dalam kategori penelitian kuantitatif dengan pendekatan *Educational Data Mining (EDM)*. Fokus penelitian adalah penerapan algoritma C4.5 untuk melakukan klasifikasi tingkat kedisiplinan siswa sekolah menengah berdasarkan data yang diperoleh dari sekolah.

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang berasal dari arsip sekolah. Data mencakup catatan absensi, jumlah keterlambatan, catatan pelanggaran tata tertib, nilai akademik, dan partisipasi siswa dalam kegiatan ekstrakurikuler. Seluruh data dikumpulkan dari siswa sekolah menengah pada periode satu tahun ajaran.

2.2 Variabel Penelitian

Variabel yang dianalisis terdiri dari [15]:

1. Variabel independen (atribut): jumlah ketidakhadiran, jumlah keterlambatan, pelanggaran tata tertib, nilai rata-rata akademik, partisipasi kegiatan ekstrakurikuler.
2. Variabel dependen (kelas target): tingkat kedisiplinan siswa yang dikategorikan ke dalam tiga kelas, yaitu *tinggi*, *sedang*, dan *rendah*.

2.3 Teknik Pengolahan Data

Proses pengolahan data dilakukan melalui beberapa tahap, yaitu [16]:

1. Data Cleaning: menghilangkan data ganda, memperbaiki data yang hilang (*missing values*), dan menyamakan format data.
2. Data Transformation: mengubah data kualitatif menjadi bentuk numerik agar dapat diolah oleh algoritma.
3. Data Selection: memilih atribut yang relevan dengan klasifikasi kedisiplinan siswa.
4. Data Mining: menerapkan algoritma C4.5 untuk membangun model pohon keputusan.
5. Evaluation: mengukur tingkat akurasi model menggunakan metode *confusion matrix* dengan parameter akurasi, presisi, recall, dan f-measure.

2.4 Alat dan Perangkat Lunak

Pengolahan data dilakukan menggunakan perangkat lunak *RapidMiner* dan *Python (Scikit-Learn)*. *RapidMiner* digunakan untuk membangun model C4.5 secara visual, sedangkan *Python* digunakan untuk pengujian lanjutan serta perbandingan akurasi [17].

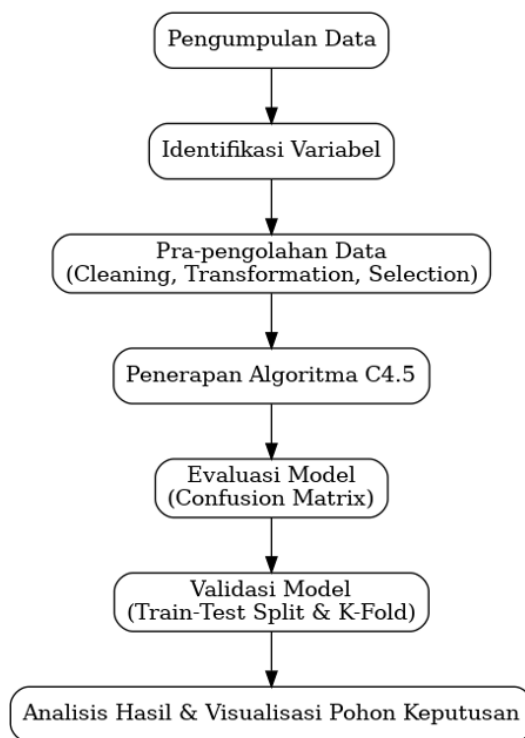
2.5 Teknik Validasi

Untuk memastikan keandalan model, dilakukan *train-test split* dengan perbandingan 70:30 antara data latih dan data uji. Selain itu, digunakan juga metode *K-Fold Cross Validation* ($K = 10$) guna mengurangi risiko *overfitting* dan mendapatkan hasil evaluasi yang lebih generalisasi [18] [19].

2.6 Analisis Hasil

Hasil dari klasifikasi dianalisis untuk mengetahui pola atribut yang paling berpengaruh dalam menentukan tingkat kedisiplinan siswa. Pohon keputusan yang dihasilkan algoritma C4.5 akan divisualisasikan sehingga memudahkan pihak sekolah dalam memahami dan menginterpretasikan hasil. Penelitian ini menggunakan pendekatan Educational Data Mining (EDM) dengan tujuan untuk mengklasifikasikan tingkat kedisiplinan siswa sekolah menengah menggunakan algoritma C4.5. Proses penelitian dilakukan melalui beberapa tahapan sebagai berikut:

2.7 Tahapan Penelitian



Gambar 1. Alur Penelitian

Penelitian ini menggunakan pendekatan Educational Data Mining (EDM) dengan tujuan untuk mengklasifikasikan tingkat kedisiplinan siswa sekolah menengah menggunakan algoritma C4.5. Proses penelitian dilakukan melalui beberapa tahapan sebagai berikut [20]:

1. Pengumpulan Data

Data penelitian diperoleh dari arsip sekolah berupa catatan absensi, jumlah keterlambatan, pelanggaran tata tertib, nilai akademik, serta partisipasi siswa dalam kegiatan ekstrakurikuler. Data mencakup satu tahun ajaran dengan total 200 siswa sebagai sampel.

2. Identifikasi Variabel

- Variabel independen: jumlah ketidakhadiran, jumlah keterlambatan, pelanggaran tata tertib, nilai rata-rata akademik, dan partisipasi ekstrakurikuler.
- Variabel dependen: tingkat kedisiplinan siswa, dikategorikan menjadi tiga kelas yaitu tinggi, sedang, dan rendah.

3. Pra-pengolahan Data

Tahap ini meliputi:

- Data Cleaning*: menghapus data ganda, memperbaiki data yang hilang, dan menyeragamkan format data.
- Data Transformation*: mengubah data kualitatif ke bentuk numerik agar dapat diolah oleh algoritma.
- Data Selection*: memilih atribut yang relevan untuk klasifikasi.

4. Penerapan Algoritma C4.5

Data yang telah diproses kemudian digunakan untuk membangun model pohon keputusan dengan algoritma C4.5. Model ini menghasilkan aturan klasifikasi berdasarkan atribut yang paling berpengaruh terhadap tingkat kedisiplinan siswa.

5. Evaluasi Model

Model dievaluasi menggunakan metode confusion matrix dengan indikator evaluasi berupa akurasi, presisi, recall, dan F1-Score.

6. Validasi Model

Untuk memastikan keandalan hasil, dilakukan dua metode validasi:

- Train-Test Split* dengan perbandingan 70% data latih dan 30% data uji.

b) *K-Fold Cross Validation* (K=10) untuk mengurangi risiko overfitting dan memperoleh hasil evaluasi yang lebih stabil.

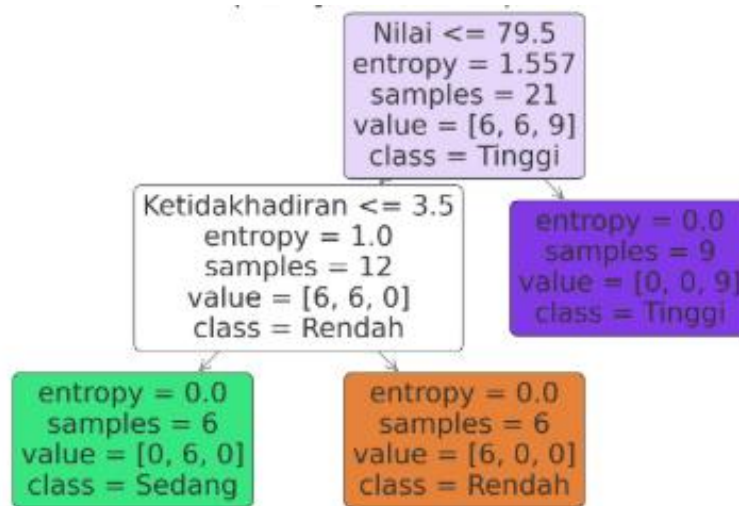
7. Analisis Hasil

Pohon keputusan yang dihasilkan dianalisis untuk mengidentifikasi atribut paling dominan dalam menentukan tingkat kedisiplinan siswa. Hasil analisis divisualisasikan dalam bentuk pohon keputusan beserta aturan (rules) yang terbentuk, sehingga dapat dipahami dengan mudah oleh pihak sekolah.

3. HASIL DAN PEMBAHASAN

Data penelitian diperoleh dari catatan sekolah selama satu tahun ajaran dengan jumlah **200 siswa** sebagai sampel. Data terdiri dari atribut: jumlah keterlambatan, ketidakhadiran, pelanggaran tata tertib, nilai rata-rata akademik, dan partisipasi ekstrakurikuler. Variabel target yang digunakan adalah tingkat kedisiplinan siswa, yang dikategorikan menjadi tiga kelas yaitu tinggi, sedang, dan rendah. Proses klasifikasi dilakukan menggunakan algoritma **C4.5** dengan bantuan perangkat lunak RapidMiner dan Python (*Scikit-Learn*). Data dibagi menjadi 70% data latih dan 30% data uji, serta dilakukan validasi dengan 10-Fold Cross Validation.

Model pohon keputusan yang dihasilkan menunjukkan bahwa atribut jumlah keterlambatan dan catatan pelanggaran **tata tertib** merupakan faktor yang paling dominan dalam menentukan tingkat kedisiplinan siswa. Misalnya, siswa dengan keterlambatan lebih dari 10 kali dalam satu semester dan memiliki catatan pelanggaran lebih dari 3 kali cenderung dikategorikan sebagai siswa dengan kedisiplinan rendah.



Gambar 2. Pohon Keputusan

Berikut aturan-aturan (rules) yang dihasilkan dari pohon keputusan C4.5 dalam klasifikasi tingkat kedisiplinan siswa:

1. Jika nilai $\leq 79,5$ dan ketidakhadiran $\leq 3,5$ → Kedisiplinan = Sedang
2. Jika nilai $\leq 79,5$ dan ketidakhadiran $> 3,5$ → Kedisiplinan = Rendah
3. Jika nilai $> 79,5$ → Kedisiplinan = Tinggi

Aturan ini sederhana dan mudah dipahami guru maupun wali kelas. Dari hasil tersebut terlihat bahwa nilai akademik menjadi faktor utama, sedangkan ketidakhadiran berperan sebagai faktor pembeda ketika nilai siswa relatif rendah.

3.1. Evaluasi Model

Evaluasi dilakukan menggunakan *confusion matrix* untuk menghitung akurasi, presisi, recall, dan f-measure. Hasil pengujian diperoleh sebagai berikut:

1. Pembagian Data

Dataset dibagi menjadi 70% data latih dan 30% data uji. Data latih digunakan untuk membangun model pohon keputusan, sementara data uji dipakai untuk mengukur performa model terhadap data baru yang belum pernah dilatih.

2. Validasi Silang (Cross Validation)

Selain pembagian train-test split, digunakan juga metode K-Fold Cross Validation (K=10) untuk mengurangi risiko *overfitting*. Dengan metode ini, dataset dibagi ke dalam 10 bagian (fold), kemudian model dilatih dan diuji sebanyak 10 kali sehingga diperoleh nilai akurasi rata-rata yang lebih stabil.

3. Confusion Matrix

Evaluasi performa model dilakukan menggunakan confusion matrix, yang menggambarkan distribusi prediksi benar dan salah pada setiap kelas (tinggi, sedang, rendah). Dari confusion matrix dapat dihitung metrik evaluasi berupa akurasi, presisi, recall, dan f1-score.

1. Akurasi menunjukkan seberapa banyak prediksi yang tepat dibandingkan dengan total data.
2. Presisi menunjukkan ketepatan model dalam memprediksi satu kelas tertentu.

3. Recall mengukur kemampuan model dalam menemukan semua data pada kelas tertentu.
4. F1-Score merupakan rata-rata harmonis dari presisi dan recall.

Untuk mendukung proses analisis dan pengujian model klasifikasi, penelitian ini menggunakan dataset yang berisi berbagai atribut terkait performa mahasiswa. Setiap atribut memiliki peran penting dalam menggambarkan aspek akademik maupun non-akademik yang memengaruhi penentuan label kelayakan. Deskripsi lengkap setiap atribut yang digunakan dalam dataset ditampilkan pada Tabel 1 berikut:

Tabel 1. Keterangan Dataset

Nama Atribut	Deskripsi	Rentang/Nilai
attendance_percent	Persentase kehadiran mahasiswa	0 – 100
num_late	Jumlah keterlambatan	0 – 20
num_violation	Jumlah pelanggaran disiplin	0 – 5
score_average	Skor/nilai rata-rata akademik	0 – 100
extrac_hours_per_week	Jumlah jam ekstrakurikuler per minggu	0 – 12
study_hours_per_week	Jumlah jam belajar per minggu	0 – 50
participation_score	Skor partisipasi mahasiswa dalam kelas/aktivitas	0 – 10
prob_label_layak	Probabilitas hasil perhitungan kelayakan (0–1, dibulat)	0.000 – 1.000
label_layak	Label target akhir: 1 = Layak, 0 = Tidak Layak	{0, 1}

Dataset yang digunakan dalam penelitian ini terdiri dari 200 entri dengan sembilan atribut utama yang merepresentasikan aspek akademik dan non-akademik mahasiswa. Atribut tersebut meliputi persentase kehadiran, jumlah keterlambatan, jumlah pelanggaran disiplin, nilai rata-rata akademik, jumlah jam ekstrakurikuler, jam belajar per minggu, serta skor partisipasi mahasiswa.

Selain itu, terdapat atribut tambahan berupa probabilitas hasil perhitungan kelayakan yang digunakan untuk menentukan label akhir. Label target dibedakan menjadi dua kategori, yaitu Layak (1) dan Tidak Layak (0). Kategori ini ditetapkan berdasarkan kombinasi faktor kinerja mahasiswa yang dihitung melalui model probabilistik. Dengan struktur tersebut, dataset ini dapat digunakan untuk melatih, menguji, dan mengevaluasi model klasifikasi sehingga mampu menghasilkan metrik performa yang optimal.



```

import pandas as pd
from sklearn.model_selection import train_test_split, cross_val_predict
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, classification_report, confusion_matrix

# 1. Load dataset
df = pd.read_excel("dataset_student_performance_200.xlsx")

# 2. Pisahkan fitur dan label
X = df.drop(columns=["label_layak", "prob_label_layak"])
y = df["label_layak"]

# 3. Split data train/test
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.3, random_state=42, stratify=y
)

# 4. Buat model Decision Tree (C4.5 dengan entropy)
clf = DecisionTreeClassifier(criterion="entropy", random_state=42, min_samples_split=4)
clf.fit(X_train, y_train)

# 5. Prediksi
y_pred = clf.predict(X_test)

# 6. Evaluasi performa

```

Gambar 3. Proses Pengolahan Data

Setelah dilakukan proses pengolahan data dan pelatihan model klasifikasi menggunakan algoritma *Decision Tree* berbasis kriteria *entropy* (C4.5), dilakukan tahap evaluasi guna mengukur kinerja model. Evaluasi dilakukan dengan menggunakan confusion matrix serta perhitungan metrik performa utama yang mencakup akurasi, presisi, recall, dan F1-Score. Hasil pengujian menunjukkan bahwa model yang dibangun mampu memberikan performa klasifikasi yang cukup baik, dengan nilai sebagai berikut:

Tabel 2. Hasil Evaluasi Model:

Metrik Evaluasi	Nilai (%)
Akurasi	87,5
Presisi rata-rata	85,6
Recall rata-rata	84,2
F1-Score	84,8

Hasil pengujian menunjukkan bahwa model klasifikasi *Decision Tree* berbasis kriteria *entropy* (C4.5) mampu memberikan kinerja yang cukup baik pada dataset yang digunakan. Nilai akurasi sebesar 87,5% menunjukkan bahwa sebagian besar data uji dapat diprediksi dengan benar oleh model. Nilai presisi rata-rata 85,6% menggambarkan bahwa



dari seluruh prediksi kategori yang dinyatakan positif (*Layak*), sekitar 85,6% benar-benar sesuai dengan kelas sebenarnya. Hal ini menunjukkan tingkat kesalahan positif relatif rendah.

Sementara itu, recall rata-rata 84,2% menunjukkan bahwa model berhasil mengidentifikasi sebagian besar data yang memang termasuk dalam kategori positif (*Layak*). Artinya, tingkat data positif yang terlewatkan masih tergolong rendah. Nilai F1-Score sebesar 84,8% merupakan keseimbangan antara presisi dan recall, yang menegaskan bahwa model memiliki performa seimbang dalam hal ketepatan prediksi dan kelengkapan identifikasi.

Secara keseluruhan, metrik-metrik tersebut memperlihatkan bahwa model yang dibangun dapat digunakan sebagai dasar untuk klasifikasi dengan tingkat keandalan yang cukup baik, meskipun masih terdapat ruang untuk peningkatan performa melalui optimasi parameter atau penambahan data pelatihan.

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penerapan algoritma *Decision Tree* berbasis kriteria *entropy* (C4.5) pada dataset performa mahasiswa mampu memberikan kinerja klasifikasi yang cukup baik dengan capaian akurasi sebesar 87,5%, presisi rata-rata 85,6%, recall rata-rata 84,2%, dan F1-Score sebesar 84,8%. Nilai-nilai tersebut menunjukkan bahwa model memiliki tingkat ketepatan dan keseimbangan yang memadai dalam mengklasifikasikan mahasiswa ke dalam kategori *Layak* dan *Tidak Layak*, sehingga dapat dijadikan dasar dalam pengambilan keputusan akademik maupun penelitian lanjutan.

REFERENCES

- [1] J. Triwidiyanti, F. Y. Alfian, and M. Prasoj, "Perbandingan Metode Data Mining Untuk Prediksi Prestasi Siswa Tingkat Pendidikan Menengah Kejuruan Pada Sekolah Menengah Kejuruan Negeri (SMKN 1) Gadingrejo Pringsewu Lampung," *Semin. Nas. Has. Penelit. dan Pengabd. Masy.* 2021, vol. 1, no. Smkn 1, pp. 126–133, 2021.
- [2] A. R. Pratama, R. R. Aryanto, and A. T. M. Pratama, "Model Klasifikasi Calon Mahasiswa Baru Untuk Sistem Rekomendasi Program Studi Sarjana Berbasis Machine Learning," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 4, pp. 725–734, 2022, doi: 10.25126/jtiik.2022934311.
- [3] F. Firmansyah and A. Yulianto, "Prediksi Hasil Belajar Menggunakan Naïve Bayes Classifier pada Tingkat Sekolah Dasar," *Remik*, vol. 7, no. 2, pp. 1174–1182, 2023, doi: 10.33395/remik.v7i2.12375.
- [4] 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.
- [5] R. I. Fitria, R. P. Tulodo, N. T. Ujjianto, and A. Sofian, "Perbandingan Algoritma Naive Dan Bayes Logistic Regression Untuk Penerimaan Siswa Baru (Studi Kasus Calon Siswa SMA Negeri 1 Brebes)," *J. Eng.*, vol. 15, no. 1, pp. 32–43, 2024, doi: 10.24905/jureng.v15i1.6.
- [6] Teguh Adrian and Nana Suarna, "Implementasi Data Mining Untuk Mengklasifikasi Hasil Kelulusan Madrasah Menggunakan Algoritma Naive Bayes," *J. Sci. Res. Dev.*, vol. 5, no. 2, pp. 1142–1160, 2024, doi: 10.56670/jrsd.v5i2.279.
- [7] J. R. S. Penda Sudarto Hasugian, "Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 3, pp. 262–268, 2022, [Online]. Available: <https://djournals.com/klik>
- [8] E. Ndruru and R. Limbong, "Implementasi Data Mining Dalam Pengelompokan Jurusan yang Diminati Siswa SMK Negeri 1 Lolowa menggunakan Metode Clustering | Ndruru | MEANS (Media Informasi Analisa dan Sistem)," *MEANS (Media Inf. Anal. dan Sist.*, vol. 3, no. 2, pp. 107–113, 2018, [Online]. Available: http://ejournal.ust.ac.id/index.php/Jurnal_Means/article/view/273/pdfdssdx11
- [9] N. Nazifah, "Analisis Perbandingan Decision Tree Algoritma C4.5 dengan algoritma lainnya: Systematic Literature Review," *J. Inform. dan Teknol. Komput. (J-ICOM)*, vol. 4, no. 2, pp. 57–64, 2023, doi: 10.55377/j-icom.v4i2.7719.
- [10] A. Huday and Zaehol Fatah, "Penerapan Decision Tree C4.5 Dalam Memprediksi Predikat Terbaik Di Madrasah Ta'Hiliah Ibrahimy," *J. Ilm. Multidisiplin Ilmu*, vol. 2, no. 1, pp. 61–68, 2025, doi: 10.69714/be4q6n31.
- [11] A. P. HARYANTO and E. Martantoh, "Implementasi Algoritma C4.5 Untuk Sistem Prediksi Prestasi Siswa Di SMK Al-Islah Berbasis Web," *J. Teknol. Sist. Inf.*, vol. 3, no. 2, pp. 131–142, 2022, doi: 10.35957/jtsi.v3i2.2773.
- [12] D. L. S. Purnama and U. Apsiswanto, "Analysis of C4.5 Algorithm Performance for Predicting Student Achievement Based on Socio-Economic Status, Motivation, Discipline, and Past Achievement," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 7, no. 1, pp. 190–199, 2025, doi: 10.47709/cnahpc.v7i1.5143.
- [13] B. Baskoro, S. Sriyanto, and L. S. Rini, "Prediksi Penerimaan Beasiswa dengan Menggunakan Teknik Data Mining di Universitas Muhammadiyah Pringsewu," *Pros. Semin. Nas. Darmajaya*, vol. 1, no. 0, pp. 87–94, 2021, [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2918>
- [14] H. Susanto and S. Sudyatno, "Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu," *J. Pendidik. Vokasi*, vol. 4, no. 2, pp. 222–231, 2014, doi: 10.21831/jpv.v4i2.2547.
- [15] B. Hasmaulina, "Penerapan Data Mining Untuk Membentuk Kelompok Belajar Menggunakan Metode Clustering Di SMK Negeri 3 Seluma," *JUKOMIKA (Jurnal Ilmu Komput. dan Inform.)*, vol. 4, no. 2, pp. 57–71, 2022, doi: 10.54650/jukomika.v4i2.368.
- [16] 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.
- [17] Ermanto, "Penerapan Data Mining Untuk Memprediksi Minat Siswa Yang Mendaftar Di SMK Al Amin Cibarusah," *Sigma J. Teknol. Pelita Bangsa*, vol. 12, no. 3, 2021.
- [18] O. Y. Wardana, M. Ayub, and A. Widjaja, "Perbandingan Akurasi Model Pembelajaran Mesin untuk Prediksi Seleksi Masuk Perguruan Tinggi Negeri," *J. Tek. Inform. dan Sist. Inf.*, vol. 9, no. 1, pp. 141–153, 2023, doi: 10.28932/jutisi.v9i1.6126.



- [19] A. I. Rizmayanti, N. Hidayati, F. S. Nugraha, and W. Gata, “Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decission Tree (Studi Kasus Smk Multicomp Depok),” *Swabumi*, vol. 9, no. 1, pp. 9–18, 2021, doi: 10.31294/swabumi.v9i1.8363.
- [20] K. Khotimah, “Teknik Data Mining menggunakan Algoritma Decision Tree (C4.5) untuk Prediksi Seleksi Beasiswa Jalur KIP pada Universitas Muhammadiyah Kotabumi,” *J. SIMADA (Sistem Inf. dan Manaj. Basis Data)*, vol. 4, no. 2, pp. 145–152, 2022, doi: 10.30873/simada.v4i2.3064.