



KLASIFIKASI EFEKTIVITAS KINERJA GURU MENGGUNAKAN ALGORITMA C4.5 PADA SMA NEGERI 9 MANDAU

Mila Hastati¹, Rahmaddeni², Yurifa Ramadhani³, Denok Wulandari⁴, Salmira⁵

^{1,2,3,5} Prodi Teknik Informatika, ⁴ Prodi Sistem Informasi

¹²³⁵ Universitas Sains dan Teknologi Indonesia, ⁴ Institut Teknologi dan Bisnis Indobaru Nasional

¹²³⁵ Jl. Purwodadi Indah KM.10, Kel. Sialangmunggu, Kec. Tuah Madani, Kota Pekanbaru, 28299

⁴ Komplek Mitra Mas Blok C 11 - 18, Tlk. Tering, Kec. Batam Kota, Kota Batam, Kep. Riau, 29444

e-mail : 2417052802124@usti.ac.id¹, rahmaddeni@usti.ac.id², 2110031802081@sar.ac.id³,

denokwulandari18@gmail.com⁴, 2417052802123@usti.ac.id⁵

ABSTRAK

Kinerja guru berperan penting dalam menentukan kualitas pendidikan. Oleh karena itu, evaluasi yang akurat dan objektif terhadap kinerja guru diperlukan untuk meningkatkan profesionalisme dan efektivitas pembelajaran. Penelitian ini bertujuan untuk mengklasifikasikan tingkat efektivitas kinerja guru di SMA Negeri 9 Mandau menggunakan algoritma C4.5 yang diimplementasikan dalam perangkat lunak RapidMiner. Data yang digunakan sebanyak 50 entri, mencakup indikator seperti kompetensi pedagogik, kehadiran, penguasaan materi, penggunaan media pembelajaran, dan hasil evaluasi siswa. Model klasifikasi dikembangkan melalui proses pelatihan dan pengujian dengan validasi silang, menghasilkan akurasi sebesar 62%, precision rata-rata 56,7%, recall 48,89%, dan F1-score 73,2%. Hasil klasifikasi menunjukkan bahwa 50% guru termasuk dalam kategori Efektif, 66,67% Cukup Efektif, dan 50% Tidak Efektif. Atribut seperti Evaluasi Atasan, Kepuasan Siswa, dan Hasil Observasi Kelas memiliki nilai Gain Ratio tertinggi, sehingga berperan signifikan dalam pembentukan struktur pohon keputusan. Temuan ini menunjukkan bahwa algoritma C4.5 mampu mengolah data evaluasi kinerja guru secara efektif dan dapat dijadikan dasar dalam pengambilan keputusan untuk program peningkatan mutu pendidikan yang lebih terarah dan berbasis data.

Kata kunci : kinerja guru, evaluasi, data mining, algoritma c4.5, rapidminer, klasifikasi

ABSTRACT

Teacher performance plays a crucial role in determining the quality of education. Therefore, accurate and objective evaluation of teacher performance is necessary to enhance professionalism and the effectiveness of learning. This study aims to classify the level of teacher performance effectiveness at SMA Negeri 9 Mandau using the C4.5 algorithm implemented in RapidMiner software. The dataset consists of 50 entries, covering indicators such as pedagogical competence, attendance, subject mastery, use of instructional media, and student evaluation results. The classification model was developed through training and testing processes with cross-validation, resulting in an accuracy of 62%, an average precision of 56.7%, recall of 48.89%, and an F1-score of 73.2%. The classification results show that 50% of teachers fall into the Effective category, 66.67% into the Quite Effective category, and 50% into the Ineffective category. Attributes such as Supervisor Evaluation, Student Satisfaction, and Classroom Observation Results have the highest Gain Ratio values, thus playing a significant role in the decision tree structure. These findings indicate that the C4.5 algorithm can effectively process teacher performance evaluation data and serve as a basis for more targeted, data-driven educational quality improvement programs.

Keywords: teacher performance, evaluation, data mining, c4.5 algorithm, rapidminer, classification.



1. PENDAHULUAN

Kualitas pendidikan sangat dipengaruhi oleh kinerja guru sebagai ujung tombak dalam proses pembelajaran. Guru yang memiliki kinerja baik mampu menciptakan suasana belajar yang kondusif, meningkatkan pemahaman siswa, serta membentuk karakter yang positif. Oleh karena itu, evaluasi kinerja guru secara objektif dan sistematis menjadi kebutuhan penting dalam upaya peningkatan mutu pendidikan.

Di SMA Negeri 9 Mandau, proses evaluasi kinerja guru selama ini dilakukan secara manual melalui observasi, penilaian kepala sekolah, serta angket dari siswa. Metode ini cenderung memakan waktu, bersifat subjektif, dan belum memanfaatkan teknologi secara optimal. Akibatnya, hasil evaluasi kurang efektif dalam memberikan umpan balik yang akurat bagi peningkatan profesionalisme guru.

Seiring dengan perkembangan teknologi informasi, analisis data secara digital dapat dimanfaatkan untuk mendukung pengambilan keputusan yang lebih cepat dan akurat. Salah satu pendekatan yang potensial adalah penggunaan algoritma *data mining*, khususnya algoritma C4.5, untuk melakukan klasifikasi terhadap efektivitas kinerja guru berdasarkan berbagai indikator penilaian. Data mining adalah proses menemukan korelasi baru yang bermakna, pola dan tren dengan memilah-milah sejumlah besar data yang tersimpan dalam repositori, menggunakan teknologi penalaran pola serta teknik-teknik statistik dan matematika (Fitriani, 2020). Algoritma C4.5 dikenal karena kemampuannya membentuk pohon keputusan yang mudah dipahami dan memiliki tingkat akurasi yang baik dalam proses klasifikasi. Dengan bantuan perangkat lunak RapidMiner, proses pengolahan dan analisis data dapat dilakukan secara efisien dan sistematis. Menurut S.E. Goodman dan S.T. Hedetniemi menyatakan bahwa algoritma adalah suatu urutan tak terbatas yang dihasilkan oleh operasi yang sudah terdefinisi dengan baik, yang masing-masing hanya memiliki waktu dan penyimpanan yang terbatas. (Rahmaddeni et al., 2024)

Beberapa penelitian sebelumnya telah menunjukkan efektivitas algoritma C4.5 dalam klasifikasi di bidang pendidikan dan kesehatan. Misalnya, penelitian oleh Hidayatuloh et al. (Hidayatuloh et al., 2022) menerapkan algoritma C4.5 untuk mengklasifikasikan capaian pembelajaran daring di MAN 3 Blitar, dan menunjukkan bahwa atribut Menciptakan memiliki nilai gain tertinggi, dengan akurasi klasifikasi mencapai 88%, termasuk dalam kategori *Good Classification*.

Penelitian lain oleh Komalasari et al. (Komalasari et al., 2024) mengevaluasi tingkat kepuasan masyarakat terhadap penggunaan aplikasi *E-Open* di Kelurahan Jatimakmur, Kota Bekasi, menggunakan algoritma C4.5 dan perangkat lunak RapidMiner versi 10.1. Hasil analisis menunjukkan tingkat akurasi klasifikasi sebesar 91,33% dan 90,67% pada perhitungan manual, dengan hasil klasifikasi berada dalam kategori Baik Sekali.

Selain itu, Gunawan et al. (Gunawan et al., 2023) menerapkan algoritma *decision tree* C4.5 untuk mengklasifikasikan data penyakit jantung dengan bantuan RapidMiner versi 9.10. Melalui proses *pre-processing*, pelatihan, pengujian, dan evaluasi menggunakan *confusion matrix*, penelitian ini menghasilkan akurasi sebesar 80,43% dan nilai AUC sebesar 0,798, menunjukkan kinerja model yang cukup baik.

Penelitian ini bertujuan untuk mengklasifikasikan efektivitas kinerja guru di SMA Negeri 9 Mandau menggunakan algoritma C4.5 berbasis RapidMiner, berdasarkan sejumlah indikator seperti kompetensi pedagogik, kehadiran, penguasaan materi, penggunaan media pembelajaran, dan hasil evaluasi siswa. Melalui pendekatan ini, diharapkan diperoleh model evaluasi yang objektif dan berbasis data, yang dapat menghasilkan pohon keputusan, aturan-aturan klasifikasi, serta nilai evaluasi seperti akurasi, *precision*, dan *recall*. Hasil penelitian ini diharapkan menjadi alat bantu evaluasi yang efektif sekaligus dasar dalam merancang program peningkatan mutu guru secara terarah.

Dalam upaya membangun sistem evaluasi yang akurat dan sistematis, pendekatan data mining menjadi salah satu solusi yang relevan. Data mining merupakan suatu proses di mana satu atau lebih sistem pembelajaran komputer digunakan untuk secara otomatis memperoleh serta menganalisis informasi, atau berupa serangkaian prosedur yang dirancang untuk menemukan nilai atau pola baru dari kumpulan data yang sebelumnya belum diketahui. (Moh. Roghib, Nining Rahaningsih, 2024). Salah satu algoritma yang banyak digunakan dalam proses data mining adalah algoritma C4.5. Algoritma C4.5 merupakan salah satu teknik yang digunakan dalam pembuatan pohon keputusan (*decision tree*). Pohon keputusan membantu menyederhanakan kumpulan data kompleks menjadi bentuk yang lebih mudah dimengerti dan digunakan dalam proses pengambilan keputusan. Tujuan utamanya adalah menemukan solusi dari suatu permasalahan dengan cara yang terstruktur (Komalasari et al.,



2024).

Algoritma C4.5 dikembangkan oleh Quinlan pada tahun 1996 sebagai penyempurnaan dari algoritma ID3. Kelemahan utama ID3 adalah hanya mampu menangani data kategorikal (nominal dan ordinal), sedangkan data numerik (interval dan rasio) belum dapat diolah (Febiani et al., 2022). Secara garis besar, proses pembuatan pohon keputusan menggunakan algoritma C4.5 mencakup tahapan berikut (Aldi Saputra et al., 2023): 1. Menentukan atribut yang akan menjadi akar pohon, 2. Membentuk cabang untuk setiap kemungkinan nilai atribut, 3. Mengelompokkan data berdasarkan masing-masing cabang. Rumus algoritma C4.5 dibagi menjadi 2 rumus yaitu mencari nilai *entropy* dan untuk mencari nilai *gain*. Berikut adalah rumus algoritma C4.5 (Komalasari et al., 2024) :

1. Rumus untuk mencari nilai *entropy* pada algoritma C4.5.

Entropi(S) adalah perkiraan jumlah bit yang diperlukan untuk membedakan suatu kelas (atau -) dari data acak dalam ruang sampel S. Entropi juga dianggap sebagai bit yang diperlukan untuk menunjukkan suatu kelas. Semakin kecil nilai entropi, semakin banyak digunakan untuk pemisahan kelas (Laturrizqi & Nurdiawan, 2023).

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \quad [1]$$

Keterangan:

S adalah Himpunan kasus (dataset)

N adalah banyaknya partisi S

Pi adalah Proporsi dari Si terhadap S

2. Rumus untuk mencari nilai *gain* pada algoritma C4.5.

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{S_i}{S} \times \text{Entropy}(S_i) \quad [2]$$

Keterangan:

S adalah Himpunan kasus (dataset)

N adalah banyaknya partisi S

Pi adalah Proporsi dari Si terhadap S

Rumus untuk mencari nilai *gain* pada algoritma C4.5.

Salah satu teknik penting dalam data mining adalah *klasifikasi*, yaitu proses mengelompokkan data ke dalam kelas-kelas tertentu berdasarkan data latih yang telah diberi label sebelumnya. Klasifikasi yaitu proses mengelompokkan objek-objek dengan karakteristik yang mirip dalam beberapa kelas. Pada umumnya pengklasifikasian dokumen diwakili oleh kalimat-kalimat penting dengan menentukan ciri-ciri atau

karakteristik (Septhya et al., 2023).

Klasifikasi bertujuan untuk mengenali pola dalam data dan membangun model yang dapat digunakan dalam pengambilan keputusan (Sidik et al., 2020). Metode klasifikasi yang paling populer salah satunya adalah *pohon keputusan* (*decision tree*). Pohon keputusan merupakan representasi dari aturan-aturan klasifikasi dalam bentuk struktur pohon yang mudah dipahami dalam bahasa alami. Metode ini mampu mengubah data kompleks menjadi aturan yang lebih sederhana dan dapat diinterpretasikan dengan baik (Suwandi, Endah Gustami, Kurnia Puji Astuti, 2022). Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda (Lumban Gaol et al., 2021).

Untuk membantu proses analisis dan visualisasi model klasifikasi, penelitian ini menggunakan *RapidMiner*, yaitu perangkat lunak open source berbasis Java yang mendukung berbagai teknik data mining, baik untuk analisis deskriptif maupun prediktif (Muzakki et al., 2023). Dengan fitur antarmuka visual yang intuitif, *RapidMiner* memungkinkan peneliti membangun model klasifikasi secara efisien tanpa perlu menulis kode pemrograman secara manual.

Pemilihan kinerja guru sebagai objek penelitian tidak terlepas dari perannya yang krusial dalam menentukan kualitas pembelajaran di sekolah. Kinerja merupakan hasil dari pelaksanaan tugas serta perilaku kerja individu yang menunjukkan tingkat pencapaian terhadap tanggung jawab yang diemban dalam periode tertentu. Oleh karena itu, penilaian kinerja perlu dilakukan secara objektif dengan menggunakan indikator-indikator yang terukur dan relevan. Tanpa alat ukur yang tepat, evaluasi terhadap kinerja guru akan sulit dilakukan secara sistematis (Priya Adhitia & Riki Winanjaya, 2023).

Dengan mengintegrasikan pendekatan data mining, metode klasifikasi pohon keputusan, dan pemanfaatan perangkat lunak *RapidMiner*, penelitian ini berupaya menghadirkan model evaluasi kinerja guru yang komprehensif, transparan, dan berbasis data.

menentukan tujuan dari penelitian, mempelajari literatur, bagaimana teknik pengumpulan data, bagaimana teknik mengolah data dan implementasi Algoritma C4.5 serta pengujian hasil (Eirlangga & Syaputra, 2022). Adapun tahapan

2. METODE PENELITIAN

Metodologi penelitian adalah tata cara melaksanakan

penelitian ini, seperti mendefinisikan ruang lingkup masalah yang ada, menganalisa masalah,



lengkap dalam metodologi penelitian ini dijelaskan sebagai berikut:

a. Jenis Penelitian

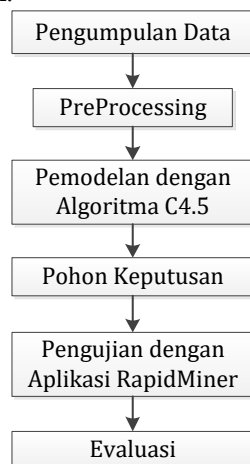
Penelitian ini merupakan penelitian kuantitatif dengan pendekatan data mining untuk melakukan klasifikasi efektivitas kinerja guru menggunakan algoritma C4.5. Penelitian dilakukan dengan bantuan perangkat lunak RapidMiner.

b. Lokasi dan Waktu Penelitian

Penelitian dilakukan di SMA Negeri 9 Mandau yang berlokasi di Jl. Stadion, Duri, Kabupaten Bengkalis, Provinsi Riau. Waktu pelaksanaan penelitian berlangsung dari bulan Mei sampai Juni 2025.

c. Tahapan Penelitian

Penelitian ini melibatkan serangkaian tahapan yang mencakup seluruh aktivitas yang dilakukan untuk mencapai tujuan yang diharapkan. Gambar berikut menyajikan alur proses dalam pelaksanaan penelitian ini.



Gambar 1. Tahapan Penelitian

Berikut adalah penjelasan tahapan penelitian berdasarkan gambar 1 :

a. Pengumpulan Data

Tahap ini merupakan proses pengumpulan data yang berkaitan dengan kinerja guru di SMA Negeri 9 Mandau. Data diperoleh dari berbagai sumber, seperti laporan penilaian kinerja guru, data kehadiran, hasil observasi kelas, evaluasi dari atasan, serta angket kepuasan siswa. Tujuan dari tahap ini adalah untuk memperoleh informasi yang akurat dan relevan sebagai dasar dalam proses klasifikasi efektivitas kinerja guru.

b. Preprocessing

Setelah data dikumpulkan, dilakukan proses preprocessing. Data Preparation mencakup semua kegiatan untuk membangun dataset yang akan dimasukkan ke dalam tools pemodelan dari data mentah /awal (Purwanto et al., 2020). Tahapan ini mencakup pembersihan data (*data cleaning*),

transformasi data, dan seleksi atribut. Proses *cleaning* adalah Proses dimana data mentah dilakukan untuk menghapus data- data yang bernilai *null* atau kosong dengan melakukan *cleaning* untuk menghapus *noise*, *missing value*, dan data yang tidak konsisten (Widiastuti et al., 2022). Proses ini bertujuan untuk memastikan bahwa data berada dalam format yang sesuai untuk dianalisis serta menghilangkan data yang tidak konsisten, redundan, atau tidak lengkap, sehingga model yang dibangun nantinya dapat bekerja secara optimal.

c. Pemodelan dengan Algoritma C4.5

Tahapan selanjutnya adalah membangun model klasifikasi menggunakan algoritma C4.5. Algoritma ini digunakan karena mampu membentuk struktur pohon keputusan berdasarkan atribut-atribut yang memiliki nilai informasi tertinggi. Model ini akan digunakan untuk mengklasifikasikan efektivitas kinerja guru ke dalam 3 (tiga) kategori, "*efektif*", "*cukup efektif*", atau "*tidak efektif*".

d. Pohon Keputusan

Hasil dari pemodelan dengan algoritma C4.5 divisualisasikan dalam bentuk pohon keputusan. Pohon ini menunjukkan jalur pengambilan keputusan berdasarkan atribut-atribut yang digunakan dalam klasifikasi. Setiap cabang pada pohon mencerminkan kondisi tertentu yang mengarah pada kesimpulan terhadap tingkat efektivitas kinerja guru.

e. Pengujian dengan Aplikasi RapidMiner

Model pohon keputusan kemudian diuji menggunakan aplikasi RapidMiner. RapidMiner merupakan platform perangkat lunak *data science* yang menyediakan lingkungan untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*). Aplikasi ini digunakan untuk bisnis, penelitian, pendidikan, pelatihan, pembuatan prototipe, dan pengembangan aplikasi. RapidMiner mendukung semua tahapan pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan (Ubaedi & Djaksana, 2022). RapidMiner digunakan untuk melakukan visualisasi dan analisis data yang baik, serta mendukung algoritma C4.5. Pengujian dilakukan dengan membagi data menjadi data pelatihan dan data pengujian untuk mengukur kinerja model.

f. Evaluasi

Tahap akhir adalah evaluasi model klasifikasi yang telah dibangun. Evaluasi dilakukan dengan mengukur *akurasi*, *presisi*, *recall*, dan *F-measure* dari model. Hasil evaluasi ini digunakan untuk menentukan seberapa baik model dalam mengklasifikasikan efektivitas kinerja guru dan



murid.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Penelitian ini bertujuan untuk mengklasifikasikan efektivitas kinerja guru di SMA Negeri 9 Mandau menggunakan algoritma C4.5 berbasis aplikasi RapidMiner. Adapun hasil penelitian diperoleh melalui beberapa tahapan sebagai berikut:

3.2 Preprocessing

Data yang digunakan pada penelitian ini adalah data kinerja guru. Data yang diperoleh peneliti dari pembagian kinerja guru SMAN 9 Mandau sebanyak 50 data. Data yang sudah diperoleh kemudian diolah menggunakan algoritma C4.5 untuk klasifikasi penilaian kinerja guru, dari hasil klasifikasi tersebut nantinya akan diuji validitasnya seberapa baik algoritma ini untuk melakukan klasifikasi data tersebut. Data penilaian kinerja guru meliputi 8 atribut antara lain yaitu kode guru, pengalaman mengajar, pelatihan terakhir, jam mengajar/minggu, hasil observasi kelas, kepuasan siswa, evaluasi atasan, Efektivitas Mengajar. Berikut data kinerja guru yang sudah dimasukkan pada Microsoft Excel:

Tabel 1. Data Kinerja Guru SMA Negeri Mandau

ID Guru	Pengalaman Mengajar (thn)	Pelatihan Terakhir (thn)	Jam Mengajar/minggu	Hasil Observasi Kelas (0-100)	Kepuasan Siswa (0-100)	Evaluasi Atasan (0-100)	Efektivitas Mengajar
1	11	0	21	76	51	50	Tidak Efektif
2	21	2	23	97	77	64	Cukup Efektif
3	24	5	31	53	97	91	Efektif
4	4	4	29	61	94	77	Cukup Efektif
5	20	4	35	92	74	74	Efektif
6	14	3	23	62	59	94	Cukup Efektif
7	24	5	34	95	94	73	Efektif
8	24	1	25	86	74	70	Cukup Efektif
9	28	5	18	93	60	65	Cukup Efektif
10	23	1	21	51	53	81	Tidak Efektif
...	...	...	...	...	...	...	...
50	16	1	25	83	71	65	Cukup Efektif

Berdasarkan tabel 1 diatas, maka dapat disimpulkan jumlah data untuk masing-masing kategori efektivitas mengajar: “Efektif” = 9 data, “Cukup Efektif”= 32 data, “Tidak Efektif “ = 9 data.

3.3 Perhitungan Menggunakan Algoritma C4.5

Perhitungan algoritma C4.5 diawali dengan menentukan atribut utama sebagai akar dengan cara menghitung jumlah semua kasus untuk kategori efektif, cukup efektif dan tidak efektif.

a) Langkah 1: Hitung Entropy Total

$$\text{Entropy}(S) = -\sum p_i \log_2(p_i) \quad [3]$$

Diketahui:

- o Jumlah total data = 50
- o Efektif = 9 $\rightarrow p_1 = 9/50 = 0.18$
- o Cukup Efektif = 32 $\rightarrow p_2 = 32/50 = 0.64$
- o Tidak Efektif = 9 $\rightarrow p_3 = 9/50 = 0.18$

$$\text{Entropy}(S) = -\left(\frac{9}{50} \log_2 \frac{9}{50} + \frac{32}{50} \log_2 \frac{32}{50} + \frac{9}{50} \log_2 \frac{9}{50}\right)$$

$$\text{Entropy}(S) = -(0.18 \times \log_2(0.18) + 0.64 \times \log_2(0.64) + 0.18 \times \log_2(0.18)) = 1.302$$

b) Langkah 2: kelompokkan data pada atribut pengalaman mengajar ke dalam 3 (tiga) kategori : Low: 0–10 tahun, Medium: 11–20 tahun, High: >20 tahun. Kemudian hitung jumlah data dalam masing-masing kelompok.

Tabel 2. Pengelompokan Atribut Pengalaman Mengajar

Kategori	Total	Efektif	Cukup Efektif	Tidak Efektif
Low	20	3	12	5
Medium	18	4	12	2
High	12	2	8	2

b.1 Berdasarkan data yang telah dikelompokkan maka hitung nilai entropy berdasarkan tiap kategorinya :

Diketahui:

- o Jumlah kategori Low = 20
- o Efektif = 3 $\rightarrow p_1 = 3/20 = 0.18$
- o Cukup Efektif = 12 $\rightarrow p_2 = 12/20 = 0.64$
- o Tidak Efektif = 5 $\rightarrow p_3 = 5/20 = 0.18$

1. Entropy (Low) :

$$\begin{aligned} &= -\left(\frac{3}{20} \log_2 \frac{3}{20} + \frac{12}{20} \log_2 \frac{12}{20} + \frac{5}{20} \log_2 \frac{5}{20}\right) = \\ &= -(0.15 \times -2.737 + 0.6 \times -0.737 + 0.25 \times -2.0) \\ &= 0.411 + 0.442 + 0.5 \\ &= 1.353 \end{aligned}$$

2. Entropy (Medium) :

$$\begin{aligned} &= -\left(\frac{4}{18} \log_2 \frac{4}{18} + \frac{12}{18} \log_2 \frac{12}{18} + \frac{2}{18} \log_2 \frac{2}{18}\right) \\ &= -(0.222 \times -2.17 + 0.667 \times -0.585 + 0.111 \times -3.17) \end{aligned}$$



$$= 0.482 + 0.390 + 0.352$$

$$= 1.224$$

3. Entropy (High) :

$$= -\left(\frac{2}{12} \log_2 \frac{2}{12} + \frac{8}{12} \log_2 \frac{8}{12} + \frac{2}{12} \log_2 \frac{2}{12}\right)$$

$$= 2 \times (0.167 \times -2.585) + 0.667 \times -0.585$$

$$= 0.863 + 0.390$$

$$= 1.253$$

b.2 Hitung Information Gain untuk menentukan selisih antara entropy total dan entropy setelah atribut digunakan :

$$\begin{aligned} \text{Gain} &= \text{Entropy (S)} - \sum \left(\frac{S_v}{S} \times \text{Entropy (S}_v)\right) \\ &= 1.302 - \left(\frac{20}{50} \times 1.353 + \frac{18}{50} \times 1.224 + \frac{12}{50} \times 1.253\right) \\ &= 1.302 - (0.541 + 0.441 + 0.301) \\ &= 1.302 - 1.283 = 0.019 \end{aligned}$$

b.3 Hitung Split Info dengan cara :

$$\begin{aligned} \text{Split Info} &= -\left(\frac{20}{50} \log_2 \frac{20}{50} + \frac{18}{50} \log_2 \frac{18}{50} + \frac{12}{50} \log_2 \frac{12}{50}\right) \\ &= -(0.4 \times -1.322 + 0.36 \times -1.473 + 0.24 \times -2.058) \\ &= 0.529 + 0.530 + 0.494 = 1.553 \end{aligned}$$

b.4 Hitung Gain Ratio, dengan cara :

$$\text{Gain Ratio} = \frac{\text{Gain}}{\text{Split Info}}$$

$$\text{Gain Ratio} = \frac{0.019}{1.553} = 0.012$$

c) Langkah 3 : Lanjutkan perhitungan pada atribut pelatihan terakhir dengan membagi menjadi 3 kategori : Low: 0–1 tahun, Medium: 2–3 tahun, High: 4–5 tahun. Sehingga datanya dapat didistribusikan menjadi :

Tabel 3. Pengelompokan Atribut Pelatihan Terakhir

Kategori	Total	Efektif	Cukup Efektif	Tidak Efektif
Low (0-1)	18	2	11	5
Medium (2-3)	17	3	12	2
High (4-5)	15	4	9	2

c.1 Berdasarkan data yang telah dikelompokkan maka hitung nilai *entropy* berdasarkan tiap kategorinya :

1. Entropy (Low) :

$$\begin{aligned} &= -\left(\frac{2}{18} \log_2 \frac{2}{18} + \frac{11}{18} \log_2 \frac{11}{18} + \frac{5}{18} \log_2 \frac{5}{18}\right) \\ &= 0.111 \times 3.17 + 0.611 \times 0.711 + 0.278 \times 1.85 \\ &= 0.352 + 0.435 + 0.514 \\ &= 1.301 \end{aligned}$$

2. Entropy (Medium) :

$$= -\left(\frac{3}{17} \log_2 \frac{3}{17} + \frac{12}{17} \log_2 \frac{12}{17} + \frac{2}{17} \log_2 \frac{2}{17}\right) =$$

$$= 0.176 \times 2.5 + 0.706 \times 0.5 + 0.118 \times 3.08$$

$$= 0.44 + 0.353 + 0.363$$

$$= 1.156$$

3. Entropy (High) :

$$= -\left(\frac{4}{15} \log_2 \frac{4}{15} + \frac{9}{15} \log_2 \frac{9}{15} + \frac{2}{15} \log_2 \frac{2}{15}\right) =$$

$$= 0.267 \times 1.91 + 0.6 \times 0.737 + 0.133 \times 2.91$$

$$= 0.51 + 0.442 + 0.387$$

$$= 1.339$$

c.2 Hitung Information Gain untuk menentukan selisih antara entropy total dan entropy setelah atribut digunakan :

$$\begin{aligned} \text{Gain} &= \text{Entropy (S)} - \sum \left(\frac{S_v}{S} \times \text{Entropy (S}_v)\right) \\ &= 1.302 - \left(\frac{18}{50} \times 1.301 + \frac{17}{50} \times 1.156 + \frac{15}{50} \times 1.339\right) \\ &= 1.302 - (0.468 + 0.393 + 0.402) \\ &= 1.302 - 1.263 \\ &= 0.039 \end{aligned}$$

c.3 Hitung Split Info dengan cara :

$$\begin{aligned} \text{Split Info} &= -\left(\frac{18}{50} \log_2 \frac{18}{50} + \frac{17}{50} \log_2 \frac{17}{50} + \frac{15}{50} \log_2 \frac{15}{50}\right) \\ &= -(-0.530 - 0.528 - 0.521) \\ &= 1.579 \end{aligned}$$

c.4 Hitung Gain Ratio, dengan cara :

$$\begin{aligned} \text{Gain Ratio} &= \frac{\text{Gain}}{\text{Split Info}} \\ \text{Gain Ratio} &= \frac{0.039}{1.579} = 0.025 \end{aligned}$$

d) Langkah 4: kelompokkan data pada atribut Jam Mengajar ke dalam 3 (tiga) kategori : Low ≤ 22 , Medium 23–30, High > 30 , sehingga datanya dapat didistribusikan sebagai berikut :

Tabel 4. Pengelompokan Atribut Jam Mengajar

Kategori	Total	Efektif	Cukup Efektif	Tidak Efektif
Low	15	2	9	4
Medium	22	4	15	3
High	13	3	8	2

Berdasarkan data yang telah dikelompokkan maka hitung nilai *entropy* berdasarkan tiap kategorinya :

1. Entropy (Low) :

$$= -\left(\frac{2}{15} \log_2 \frac{2}{15} + \frac{9}{15} \log_2 \frac{9}{15} + \frac{4}{15} \log_2 \frac{4}{15}\right) = 1.339$$

2. Entropy (Medium) :

$$= -\left(\frac{4}{22} \log_2 \frac{4}{22} + \frac{15}{22} \log_2 \frac{15}{22} + \frac{3}{22} \log_2 \frac{3}{22}\right) = 1.214$$

3. Entropy (High) :

$$= -\left(\frac{3}{13} \log_2 \frac{3}{13} + \frac{8}{13} \log_2 \frac{8}{13} + \frac{2}{13} \log_2 \frac{2}{13}\right) =$$



1.336

- d.2. Hitung Information Gain untuk menentukan selisih antara entropy total dan entropy setelah atribut digunakan :

$$\begin{aligned} \text{Gain} &= \text{Entropy}(S) - \sum \left(\frac{k_v}{S} \times \text{Entropy}(S_v) \right) \\ &= 1.302 - \left(\frac{15}{50} \times 1.339 + \frac{22}{50} \times 1.214 + \frac{13}{50} \times 1.336 \right) \\ &= 1.302 - (0.402 + 0.534 + 0.347) \\ &= 1.302 - 1.283 \\ &= 0.019 \end{aligned}$$

- d.3. Hitung Split Info dengan cara :

$$\begin{aligned} \text{SplitInfo} &= - \left(\frac{15}{50} \log_2 \frac{15}{50} + \frac{22}{50} \log_2 \frac{22}{50} + \frac{13}{50} \log_2 \frac{13}{50} \right) \\ &= - (0.3x - 1.737 + 0.44x - 1.184 + 0.26x - 1.943) \\ &= 0.521 + 0.521 + 0.505 \\ &= 1.547 \end{aligned}$$

- d.4. Untuk Gain Ratio, tambahkan split info, lalu:

$$\begin{aligned} \text{Gain Ratio} &= \frac{\text{Gain}}{\text{Split Info}} \\ \text{Gain Ratio} &= \frac{0.019}{1.547} = 0.012 \end{aligned}$$

- e) Langkah 5 : kelompokkan data pada atribut Observasi di Kelas ke dalam 3 (tiga) kategori : Low 50-69, Medium 70-84, High 85-100, sehingga datanya dapat didistribusikan sebagai berikut :

Tabel 5. Pengelompokan Atribut Observasi di Kelas

Kategori	Total	Efektif	Cukup Efektif	Tidak Efektif
Low	14	1	8	5
Medium	19	3	13	3
High	18	5	11	1

- e.1. Berdasarkan data yang telah dikelompokkan maka hitung nilai *entropy* berdasarkan tiap kategorinya :

1. Entropy (Low) :

$$= - \left(\frac{1}{14} \log_2 \frac{1}{14} + \frac{8}{14} \log_2 \frac{8}{14} + \frac{5}{14} \log_2 \frac{4}{14} \right) = 1.262$$

2. Entropy (Medium) :

$$= - \left(\frac{3}{19} \log_2 \frac{3}{19} + \frac{13}{19} \log_2 \frac{13}{19} + \frac{3}{19} \log_2 \frac{3}{19} \right) = 1.214$$

3. Entropy (High) :

$$= - \left(\frac{5}{18} \log_2 \frac{5}{18} + \frac{11}{18} \log_2 \frac{11}{18} + \frac{1}{18} \log_2 \frac{1}{18} \right) = 1.167$$

- e.2. Hitung Information Gain untuk menentukan selisih antara entropy total dan entropy setelah atribut digunakan :

$$\begin{aligned} \text{Gain} &= \text{Entropy}(S) - \sum \left(\frac{k_v}{S} \times \text{Entropy}(S_v) \right) \\ &= 1.302 - \left(\frac{14}{50} \times 1.262 + \frac{19}{50} \times 1.214 + \right. \end{aligned}$$

$$\begin{aligned} &\left. \frac{17}{50} \times 1.167 \right) \\ &= 1.302 - (0.354 + 0.461 + 0.397) \\ &= 1.302 - 1.212 \\ &= 0.090 \end{aligned}$$

- e.3. Hitung Split Info dengan cara :

$$\begin{aligned} \text{SplitInfo} &= - \left(\frac{14}{50} \log_2 \frac{14}{50} + \frac{19}{50} \log_2 \frac{19}{50} + \frac{17}{50} \log_2 \frac{17}{50} \right) \\ &= 0.28x1.83 + 0.38x1.395 + 0.34x1.555 \\ &= 0.512 + 0.530 + 0.529 \\ &= 1.571 \end{aligned}$$

- e.4. Untuk Gain Ratio, tambahkan split info, lalu:

$$\begin{aligned} \text{Gain Ratio} &= \frac{\text{Gain}}{\text{Split Info}} \\ \text{Gain Ratio} &= \frac{0.090}{1.571} = 0.083 \end{aligned}$$

- f) Langkah 6 : kelompokkan data pada atribut Kepuasan Siswa ke dalam 3 (tiga) kategori : Low 50-44, Medium 65-79, High 80-100, sehingga datanya dapat didistribusikan sebagai berikut :

Tabel 6. Pengelompokan Atribut Kepuasan Siswa

Kategori	Total	Efektif	Cukup Efektif	Tidak Efektif
Low	17	2	8	7
Medium	19	4	14	1
High	14	3	10	1

- f.1. Berdasarkan data yang telah dikelompokkan maka hitung nilai *entropy* berdasarkan tiap kategorinya :

1. Entropy (Low) :

$$= - \left(\frac{2}{17} \log_2 \frac{2}{17} + \frac{8}{17} \log_2 \frac{8}{17} + \frac{7}{17} \log_2 \frac{7}{17} \right) = 1.402$$

2. Entropy (Medium) :

$$= - \left(\frac{4}{19} \log_2 \frac{4}{19} + \frac{14}{19} \log_2 \frac{14}{19} + \frac{1}{19} \log_2 \frac{1}{19} \right) = 1.024$$

3. Entropy (High) :

$$= - \left(\frac{3}{14} \log_2 \frac{3}{14} + \frac{10}{14} \log_2 \frac{10}{14} + \frac{1}{14} \log_2 \frac{1}{14} \right) = 1.093$$

- f.2. Hitung *Information Gain* untuk menentukan selisih antara entropy total dan entropy setelah atribut digunakan :

$$\begin{aligned} \text{Gain} &= \text{Entropy}(S) - \sum \left(\frac{k_v}{S} \times \text{Entropy}(S_v) \right) \\ &= 1.302 - \left(\frac{17}{50} \times 1.402 + \frac{19}{50} \times 1.024 + \frac{14}{50} \times 1.093 \right) \\ &= 1.302 - (0.477 + 0.389 + 0.306) \end{aligned}$$



$$= 1.302 - 1.172$$

$$= 0.130$$

f.3. Hitung Split Info dengan cara :

$$\begin{aligned} \text{SplitInfo} &= -\left(\frac{17}{50} \log_2 \frac{17}{50} + \frac{19}{50} \log_2 \frac{19}{50} + \frac{14}{50} \log_2 \frac{14}{50}\right) \\ &= (0.34x - 1.555 + 0.38x - 1.395 + 0.28x - 1.83) \\ &= 0.528 + 0.530 + 0.512 \\ &= 1.570 \end{aligned}$$

f.4. Untuk Gain Ratio, tambahkan split info, lalu:

$$\begin{aligned} \text{Gain Ratio} &= \frac{\text{Gain}}{\text{Split Info}} \\ \text{Gain Ratio} &= \frac{0.130}{1.570} = 0,104 \end{aligned}$$

g) Langkah 7 : kelompokkan data pada atribut Evaluasi Atasan ke dalam 3 (tiga) kategori : Low 50-65, Medium 66-80, High 81-100, sehingga datanya dapat didistribusikan sebagai berikut :

Tabel 7. Pengelompokan Atribut Evaluasi Atasan

Kategori	Total	Efektif	Cukup Efektif	Tidak Efektif
Low	20	1	11	8
Medium	17	3	13	1
High	13	5	8	0

g.1. Berdasarkan data yang telah dikelompokkan maka hitung nilai *entropy* berdasarkan tiap kategorinya :

$$\begin{aligned} 1. \text{ Entropy (Low) :} \\ &= -\left(\frac{1}{20} \log_2 \frac{1}{20} + \frac{11}{20} \log_2 \frac{11}{20} + \frac{8}{20} \log_2 \frac{8}{20}\right) = \\ &= 1.217 \end{aligned}$$

$$\begin{aligned} 2. \text{ Entropy (Medium) :} \\ &= -\left(\frac{3}{17} \log_2 \frac{3}{17} + \frac{13}{17} \log_2 \frac{13}{17} + \frac{1}{17} \log_2 \frac{1}{17}\right) = \\ &= 0.975 \end{aligned}$$

$$\begin{aligned} 3. \text{ Entropy (High) :} \\ &= -\left(\frac{5}{13} \log_2 \frac{5}{13} + \frac{8}{13} \log_2 \frac{8}{13} + \frac{0}{13} \log_2 \frac{0}{13}\right) = \\ &= 0.963 \end{aligned}$$

g.2. Hitung Information Gain untuk menentukan selisih antara entropy total dan entropy setelah atribut digunakan :

$$\text{Gain} = \text{Entropy (S)} - \sum \left(\frac{S_v}{S} \times \text{Entropy (S}_v\right)$$

$$= 1.302 - \left(\frac{20}{50} \times 1.217 + \frac{17}{50} \times 0.975 + \frac{13}{50} \times 0.963\right)$$

$$\begin{aligned} &= 1.302 - (0.487 + 0.331 + 0.250) \\ &= 1.302 - 1.068 \\ &= 0.234 \end{aligned}$$

g.3. Hitung Split Info dengan cara :

$$\text{SplitInfo} = -\left(\frac{20}{50} \log_2 \frac{20}{50} + \frac{17}{50} \log_2 \frac{17}{50} + \frac{13}{50} \log_2 \frac{13}{50}\right)$$

$$\frac{13}{50} \log_2 \frac{13}{50}$$

=

$$\begin{aligned} &(0.4 \times 1.322 + 0.34 \times 1.555 + 0.26 \times 1.943) \\ &= 0.529 + 0.528 + 0.505 \\ &= 1.562 \end{aligned}$$

g.4. Untuk Gain Ratio, tambahkan split info, lalu:

$$\begin{aligned} \text{Gain Ratio} &= \frac{\text{Gain}}{\text{Split Info}} \\ \text{Gain Ratio} &= \frac{0.234}{1.562} = 0.057 \end{aligned}$$

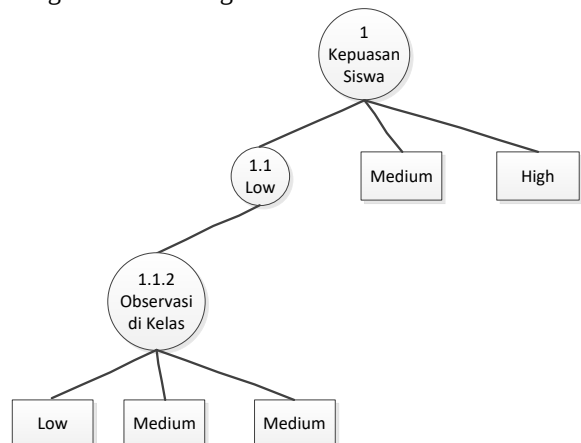
3.4 Pohon Keputusan

Decession Tree adalah sebuah diagram alir yang berbentuk seperti struktur pohon yang mana setiap internal node menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan output dari pengujian dan *leaf node* (bagian daun) menyatakan kelas-kelas atau distribusi kelas Untuk membangun pohon(Nas, 2021). keputusan C4.5 dari perhitungan di atas, maka sebaiknya menggunakan pendekatan *top-down recursive* berdasarkan Gain Ratio tertinggi. Tahap awal dalam membangun pohon keputusan adalah dengan memilih atribut dengan gain ratio tertinggi dari perhitungan sebelumnya. Berdasarkan tabel di bawah, maka gain ratio tertinggi ada pada atribut kepuasan siswa.

Tabel 8. Gain Ratio

No	Atribut	Gain Ratio
1	Pengalaman Mengajar	0.012
2	Pelatihan Terakhir	0.025
3	Jam Mengajar	0.012
4	Hasil Observasi	0.083
5	Kepuasan Siswa	0,104
6	Evaluasi Atasan	0.057

Selanjutnya struktur pohon keputusan dapat digambarkan sebagai berikut :



Gambar 2 Pohon Keputusan

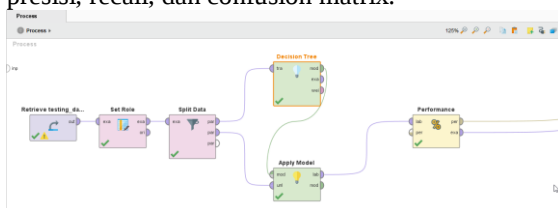


Keterangan Gambar :

Agar sebuah node dapat langsung menjadi “daun” (leaf) dalam pohon keputusan, distribusi kelasnya harus cukup dominan, misalnya: $\geq 80\%$ dari data memiliki label yang sama, atau Entropi sangat rendah (mendekati 0). Pada atribut kepuasan siswa, terdapat distribusi data sebagai berikut : 11 dari 20 pada kategori Cukup Efektif (55%), 8 dari 20 pada kategori Tidak Efektif (40%) dan entropi pada kategori Low masih tinggi bernilai 1.217. sehingga perlu displit lagi karena datanya masih terlalu beragam untuk langsung diklasifikasikan sebagai satu kelas. Untuk mengurangi entropi, maka harus mencari atribut lain yang bisa memecah data ini menjadi subset yang lebih homogen. Berdasarkan gain ratio pada tabel 2, maka nilai gain tertinggi berikutnya adalah pada pilihan atribut observasi di kelas. Gain Ratio-nya bernilai 0.083 dan efektif dalam memisahkan “Tidak Efektif” dari yang lain. Penilaian atribut Kepuasan Siswa pada kategori Medium/High didominasi oleh Cukup Efektif dan “Tidak Efektif” hanya bernilai kecil saja maka bisa disimpulkan sebagai “Cukup Efektif”. Pohon keputusan berhenti pada observasi di kelas karena: Setiap cabang setelahnya sudah cukup homogen, tidak ada informasi tambahan signifikan dari atribut lain dan pohon sudah mencapai titik klasifikasi yang cukup akurat.

3.5 Pengujian dengan Aplikasi RapidMiner

Untuk memvalidasi hasil analisis pohon keputusan yang telah dibangun secara teoritis pada bagian sebelumnya, dilakukan pengujian menggunakan RapidMiner Studio, sebuah perangkat lunak data mining berbasis GUI (Graphical User Interface) yang memudahkan proses machine learning. RapidMiner menyediakan berbagai operator yang dapat digunakan untuk membangun model pohon keputusan berbasis algoritma C4.5 (Decision Tree), melakukan pemisahan data (*data splitting*), menerapkan model, serta mengevaluasi kinerjanya melalui berbagai metrik evaluasi seperti akurasi, presisi, recall, dan confusion matrix.



Gambar 3 Pembentukan Model RapidMiner

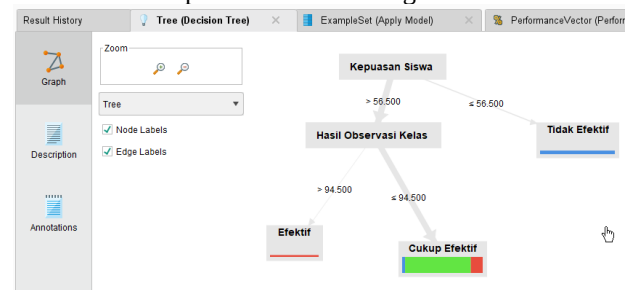
Dataset dibagi menjadi dua bagian, yaitu data pelatihan (training) dan data pengujian (testing). Proses analisis menggunakan algoritma Decision

Tree dilakukan pada data pengujian, sehingga keluaran yang dihasilkan dapat dijadikan dasar dalam interpretasi hasil model klasifikasi.

Row	Class	predicted...	confidence...	confidence...	Ratio Data	Pengamatan...	Prediction...	Actual Class...	Confusion Ma...	Confusion Ma...
1	Tidak Efektif	Tidak Efektif	1	0	0	1	11	0	21	51
2	Cukup Efektif	Efektif	0	0	1	2	21	0	23	57
3	Efektif	Cukup Efektif	0.037	0.015	0.143	3	24	0	39	53
4	Cukup Efektif	Cukup Efektif	0.037	0.015	0.143	4	4	29	61	84
5	Cukup Efektif	Cukup Efektif	0.037	0.015	0.143	6	14	3	23	62
6	Efektif	Efektif	0	0	1	7	24	0	34	65
7	Cukup Efektif	Cukup Efektif	0.037	0.015	0.143	12	26	0	27	67
8	Tidak Efektif	Cukup Efektif	0.037	0.015	0.143	13	4	1	36	68
9	Cukup Efektif	Tidak Efektif	1	0	0	15	30	1	36	64
10	Cukup Efektif	Cukup Efektif	0.037	0.015	0.143	23	19	1	29	65
11	Efektif	Cukup Efektif	0.037	0.015	0.143	26	14	4	22	61
12	Tidak Efektif	Cukup Efektif	0.037	0.015	0.143	33	19	1	33	61
13	Cukup Efektif	Cukup Efektif	0.037	0.015	0.143	36	27	3	23	66
14	Cukup Efektif	Cukup Efektif	0.037	0.015	0.143	41	28	4	31	64

Gambar 4. Hasil Pengolahan Decision Tree C45 Model RapidMiner

Model *Decision Tree* yang telah dibangun menghasilkan struktur pohon keputusan yang divisualisasikan pada Gambar 5 sebagai berikut.:



Gambar 5. Hasil Pohon Keputusan

Pada tahapan ini dilakukan pengukuran tingkat akurasi, tujuan dari mengukur akurasi adalah untuk melihat berapa nilai akurasi pada prediksi. Berikut ini adalah hasil proses prediksi untuk pengukuran tingkat aurasi.

	True Tidak Efektif	True Cukup Efektif	True Efektif	Class Precision
pred. Tidak Efektif	1	1	0	50.00%
pred. Cukup Efektif	2	8	2	66.67%
pred. Efektif	0	1	1	50.00%
class recall	33.33%	80.00%	33.33%	

Gambar 6. Nilai Accuracy

1. Pengujian Model

Data dibagi menjadi: data latih (70%) dan data uji (30%). Model pohon yang dibangun dari data latih diuji pada data uji dan dilakukan evaluasi dengan *akurasi*, *precision*, *recall*, dan *F1-score*.

a. Akurasi

Akurasi adalah salah satu metrik evaluasi paling umum dalam *machine learning* dan data mining, khususnya dalam klasifikasi. (Sadeli & Lawanda, 2023)

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi yang Benar}}{\text{Total Jumlah Data}} \quad [4]$$

$$\text{Akurasi} = \frac{31}{50} = 0,62 = 62\%$$

b. Precision (rata-rata / macro avg)

Precision adalah ukuran seberapa tepat model



ketika memprediksi suatu kelas (Sadeli & Lawanda, 2023)

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}} \quad [5]$$

- o True Positive (TP) : data benar yang diprediksi benar
- o False Positive (FP) : data salah yang diprediksi salah

Berdasarkan perhitungan data klasifikasi kinerja guru sebanyak 50 data maka :

1. Efektif:
 - o Prediksi Efektif = 2 (semuanya benar)
 - o Precision = $\frac{2}{2} = 1.0$
2. Cukup Efektif:
 - o Prediksi Cukup Efektif = 7
 - o Benar = 6
 - o Precision = $\frac{6}{7} = 0,857$
3. Tidak Efektif:
 - o Prediksi Tidak Efektif = 0
 - o Precision = 0 (*undefined* → biasanya dianggap 0)

$$\text{Macro Precision} = \frac{1.0 + 0.857 + 0.0}{3} = 0,5 \text{ atau } 50\%$$

c. Recall (rata-rata / macro avg)

Recall adalah metrik evaluasi dalam klasifikasi yang menunjukkan kemampuan model dalam menemukan semua data yang relevan untuk suatu kelas. (Sadeli & Lawanda, 2023)

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \quad [6]$$

- o True Positive (TP): Jumlah data yang benar-benar termasuk kelas tertentu dan berhasil diprediksi dengan benar.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan terkait klasifikasi efektivitas kinerja guru pada SMA Negeri 9 Mandau dengan Algoritma C4.5 maka peneliti dapat memberikan kesimpulan sebagai berikut :

1. Algoritma C4.5 mampu membentuk model klasifikasi kinerja guru dengan akurasi sebesar 62%, yang menunjukkan performa model cukup baik dalam mengelompokkan kategori efektivitas kinerja guru.
2. Nilai precision rata-rata (56,67%), recall (49,89%), dan F1-score (73,2%) menggambarkan bahwa model memiliki kemampuan yang seimbang dalam mengidentifikasi guru yang efektif dan kurang efektif, meskipun masih terdapat ruang untuk perbaikan.

- o False Negative (FN): Jumlah data yang seharusnya termasuk kelas tersebut, tapi gagal terdeteksi oleh model.

Berdasarkan data yang diuji sebanyak 50 data ada 9 data untuk kategori Efektif, 32 data untuk kategori Cukup Efektif dan 9 data untuk Tidak Efektif adalah jumlah yang diprediksi Benar untuk masing-masing kelas (alias True Positive untuk masing-masing):

Tabel 9. Recall

Kelas Efektivitas	True Positive (TP)	Jumlah Aktual	Recall
Tidak Efektif	9	7	$= \frac{9}{9+7} = 56,25\%$
Cukup Efektif	32	17	$= \frac{32}{32+17} = 65,31\%$
Efektif	9	7	$= \frac{9}{9+7} = 56,25\%$

$$\text{Rata-rata Recall} = (33,33\% + 80,00\% + 33,33\%) / 3 = 49,89\%$$

d. F1-score

F1-Score adalah metrik evaluasi yang digunakan untuk mengukur kinerja model klasifikasi, khususnya saat data tidak seimbang (misalnya jumlah antar kelas berbeda jauh). (Sadeli & Lawanda, 2023)

$$\text{Rumus : F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad [6]$$

$$\text{F1-Score} = 2 \times \frac{(0,619 \times 0,5927)}{(0,619 + 0,5927)} = 2 \times (0,366) = 0,732 \text{ atau } 73,20\%$$

3. Hasil klasifikasi menunjukkan distribusi guru sebagai berikut: 50% Efektif, 66,67% Cukup Efektif, dan 50% Tidak Efektif, mencerminkan adanya variasi signifikan dalam kualitas kinerja guru di sekolah tersebut.

5. REFERENSI

- Aldi Saputra, Triyani Arita Fitri, Karpen, & Susanti. (2023). Penerapan Data Mining Algoritma C4.5 Dalam Memprediksi Predikat Kelulusan Mahasiswa Di Politeknik Kampar. *SATIN - Sains Dan Teknologi Informasi*, 9(1), 149–157. <https://doi.org/10.33372/stn.v9i1.990>
- Eirlangga, Y. S., & Syaputra, A. E. (2022). Klasifikasi Penjurusan pada Sekolah Menengah Atas (SMA) dengan Metode Algoritma C4.5. *Jurnal Informasi Dan*



- Teknologi*, 4(3), 160–165.
<https://doi.org/10.37034/jidt.v4i3.235>
- Febiani, N., Fauzan, A. C., & Huda, M. M. (2022). Implementasi Algoritma Decision Tree C4.5 Dengan Improvisasi Mean Dan Median Pada Dataset Numerik. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 5(1), 105.
<https://doi.org/10.37600/tekinkom.v5i1.435>
- Fitriani, E. (2020). Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *Sistemasi*, 9(1), 103.
<https://doi.org/10.32520/stmsi.v9i1.596>
- Gunawan, I. M. A. O., Saraswati, I. D. A. I., Agung, I. D. G. R., & Putra, I. P. E. (2023). *Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer* (pp. 73–83). *Jurnal Teknologi dan Sistem Informasi Bisnis (JTeksis)*.
- Hidayatuloh, M. A., Kartika, K. P., & Permadi, D. F. H. (2022). Implementasi Algoritma C4.5 Untuk Memprediksi Capaian Pembelajaran Daring (Studi Kasus Siswa MAN 3 Blitar). *Jurnal Algoritme*, 3(1), 33–47.
<https://doi.org/10.35957/algoritme.v3i1.3292>
- Komalasari, Y., Puspitasari, N. R., & Chalimatusadiah, C. (2024). Penerapan Algoritma C4.5 Untuk Menentukan Kepuasan Pengguna Aplikasi E-Open Study Kasus : Kelurahan Jati Makmur. *Computer Science (CO-SCIENCE)*, 4(2), 163–172.
<https://doi.org/10.31294/coscience.v4i2.3387>
- Laturrizqi, W., & Nurdiawan, O. (2023). Implementasi Algoritma C4.5 Untuk Menentukan Penerima Bantuan Sosial. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 373–377.
<https://doi.org/10.36040/jati.v7i1.6312>
- Lumban Gaol, L. Y., Safii, M., & Suhendro, D. (2021). Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5. *Brahmana : Jurnal Penerapan Kecerdasan Buatan*, 2(2), 97–106.
<https://doi.org/10.30645/brahmana.v2i2.71>
- Moh.Roghib, Nining Rahaningsih, R. D. D. (2024). *Penerapan Algoritma C4.5 Untuk Seleksi Penjurusan Siswa Baru Pada Sekolah Menengah Kejuruan (Studi Kasus : SMK Plus Al-Hilal Arja* (pp. 861–866). *JATI (Jurnal Mahasiswa Teknik Informatika)*.
- Muzakki, F., Ubaydillah, I., Assyiami, N. R., & Soleha, S. (2023). Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Jantung Menggunakan Rapidminer. *Jurnal Komputer Antartika*, 2(2), 71–79.
<https://doi.org/10.70052/jka.v2i2.304>
- Nas, C. (2021). Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5. *Jurnal Manajemen Informatika (JAMIKA)*, 11(2), 131–145.
<https://doi.org/10.34010/jamika.v11i2.5506>
- Priya Adhitia, & Riki Winanjaya. (2023). Penerapan Data Mining Dalam Menentukan Tingkat Kepuasan Orangtua Terhadap Kinerja Guru Pada Masa Pandemi Covid-19. *Jurnal Elektronika Dan Teknik Informatika Terapan (JENTIK)*, 1(4), 201–216.
<https://doi.org/10.59061/jentik.v1i4.527>
- Purwanto, A., Primajaya, A., & Voutama, A. (2020). Penerapan Algoritma C4.5 Dalam Prediksi Potensi Tingkat Kasus Pneumonia Di Kabupaten Karawang. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 8(4), 390.
<https://doi.org/10.26418/justin.v8i4.41959>
- Rahmaddeni, Syarfi Aziz, Zairi Saputra, & Ryan Ismanizan. (2024). Implementasi Algoritma SVM dan C4.5 dalam Klasifikasi Calon Penerimaan Beasiswa. *Jurnal KomtekInfo*, 11(3), 81–88.
<https://doi.org/10.35134/komtekinfo.v11i3.524>
- Sadeli, A. F., & Lawanda, I. I. (2023). Recall, Precision, and F-Measure for Evaluating Information Retrieval System in Electronic Document Management Systems (EDMS). *Khazanah Al-Hikmah : Jurnal Ilmu Perpustakaan, Informasi, Dan Kearsipan*, 11(2), 231–241.
<https://doi.org/10.24252/kah.v11i2a8>
- Septyha, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, Irawan, Y., & Hayami, R. (2023). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 15–19.
<https://doi.org/10.57152/malcom.v3i1.591>
- Sidik, A. D. W. M., Himawan Kusumah, I., Suryana, A., Edwinanto, Artiyasa, M., & Pradiftha Junfithrana, A. (2020). Gambaran Umum Metode Klasifikasi Data Mining. *FIDELITY : Jurnal Teknik Elektro*, 2(2), 34–38.
<https://doi.org/10.52005/fidelity.v2i2.111>
- Suwandi, Endah Gustami, Kurnia Puji Astuti, Y. (2022). Klasifikasi Barang Menggunakan Algoritma C4.5 Dalam Penentuan Prediksi Stok Barang Pada PT. Aerofood Indonesia. *Reputasi : Jurnal Rekayasa Perangkat Lunak*, 3(2), 12–18.
- Ubaedi, I., & Djaksana, Y. M. (2022). Optimasi Algoritma C4.5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Kredit. *JSiI (Jurnal*



Sistem Informasi), 9(1), 17–26.
<https://doi.org/10.30656/jsii.v9i1.3505>
Widiastuti, T., Karsa, K., & Juliane, C. (2022).
Widiastuti, T., Karsa, K., & Juliane, C. (2023).
Evaluasi Tingkat Kepuasan Mahasiswa

Terhadap Pelayanan Akademik Menggunakan
Metode Klasifikasi Algoritma C4. 5.
Technomedia Journal, 7(3 Februari), 364-380.
Technomedia Journal, 7(3), 364–380.