



Penerapan Data Mining untuk Prediksi Kelulusan Siswa SD Negeri Sukaresmi Kota Bogor Dengan Alogaritma C4.5

Nurul Elvida^{1*}, Adika May Sari¹

¹ Universitas Lampung, Indonesia

 nurulelvida02@gmail.com*

Abstract

Education is currently very important for enlightening the future generation and thus must be prioritized. This study at SDN Sukaresmi Kota Bogor aims to meet graduation requirements for students who must attend classes and be present at school. This research uses the C4.5 Algorithm method to form a decision tree, which requires grade and attendance data to assess student quality. Data mining collects and classifies data using RapidMiner tools for decision-making. The decision tree process transforms data into a tree model, which is then converted into rules and simplified. Applying data mining is crucial for facilitating data management, predicting graduation, identifying problems, and analyzing data. The objectives of this research are to predict students' eligibility for graduation, discuss the decision tree using the C4.5 Algorithm, and understand how students meet graduation requirements. This research is also a requirement for graduation from the Information Systems Bachelor's program at Universitas Bina Sarana Informatika. The results of this study obtained data showing 91.00% accuracy, with 1 student achieving Excellent graduation status, 32 students with Good status, 16 students with Fairly Good status, 5 students with Fair status, and 1 student with Satisfactory status.

Keywords: Application of Data Mining, C4 Algorithm, Predict the Graduation

ARTICLE INFO

Article history:

Received

September 11,
2024

Revised

October 22, 2024

Accepted

November 09,
2024

Published by
ISSN

CV. Creative Tugu Pena
2774-7077

Website

<https://attractivejournal.com/index.php/bce/>

This is an open access article under the CC BY SA license

<https://creativecommons.org/licenses/by-sa/4.0/>



PENDAHULUAN

Pendidikan saat ini sangat penting dalam upaya mencerdaskan generasi masa depan yang lebih baik, karena itu pendidikan menjadi salah satu hal yang harus diutamakan dalam kehidupan dengan adanya pendidikan banyak ilmu yang dapat diraih dalam kehidupan. Dalam Penelitian ini dilakukan di SDN Sukaresmi Kota Bogor, dalam memenuhi syarat kelulusan siswa diharuskan mampu mengikuti pembelajaran yang diajarkan oleh guru untuk mendapatkan nilai yang sesuai dengan standar kelulusan sekolah, kehadiran siswa dalam mengikuti pembelajaran di sekolah pun menjadi salah satu syarat kelulusan sekolah.

Maka dari itu dilakukan penelitian ini, untuk memecahkan masalah dengan menggunakan teknik Alogaritma C4.5. "Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (Decision Tree)"(Maqfiroh1 & Sri Mujiyono2, 2022). Algoritma ini berfungsi untuk menganalisis data dan mengidentifikasi pola yang dapat membantu dalam pengambilan keputusan berdasarkan kriteria tertentu. Dengan pendekatan ini, data yang kompleks dapat disederhanakan menjadi bentuk yang

lebih mudah dipahami, sehingga memudahkan pemangku kepentingan dalam membuat keputusan yang tepat.

Dalam penelitian ini, langkah-langkah yang diambil diawali dengan pengumpulan data yang relevan terkait permasalahan yang ingin dipecahkan. Selanjutnya, data tersebut akan diproses dan dipersiapkan untuk membangun model pohon keputusan menggunakan algoritma C4.5.(Rofiani dkk., 2024) Proses ini melibatkan pemilihan atribut yang paling signifikan, pembagian data menjadi set pelatihan dan set pengujian, serta penerapan algoritma untuk menghasilkan pohon keputusan. Setelah model pohon keputusan terbentuk, langkah berikutnya adalah melakukan evaluasi terhadap kinerja model tersebut. Evaluasi ini penting untuk mengetahui seberapa baik algoritma C4.5 dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya.(Tri Widiastuti dkk., t.t.) Dalam penelitian ini, kami akan menggunakan beberapa metrik evaluasi, seperti akurasi, presisi, dan recall, untuk memberikan gambaran yang jelas mengenai efektivitas model. Dengan demikian, diharapkan hasil penelitian ini tidak hanya memberikan solusi terhadap permasalahan yang diangkat, tetapi juga memberikan kontribusi ilmiah dalam penggunaan algoritma C4.5 sebagai salah satu metode analisis data yang efektif dalam konteks yang lebih luas.

Membutuhkan informasi yang jelas dan akurat melibatkan data-data seperti nilai, absensi sebagai bahan yang dibutuhkan dalam menentukan kualitas siswa secara umum. Data mining merupakan serangkaian proses pengumpulan, dan dikelompokkan menjadi beberapa kelas untuk melakukan pengolahan data menggunakan tools Rapid Miner yang hasilnya dapat digunakan untuk mengambil sebuah keputusan.(Pascalina dkk., 2023). Proses pada pohon keputusan adalah mengubah suatu bentuk data (table) menjadi model pohon, merubah model pohon menjadi *rule*, dan menyederhanakan *rule*. Penerapan *data mining* merupakan element yang terpenting untuk perkembangan dalam proses pengolahan data.(Putra, 2023)

Dalam memprediksi kelulusan yang dimana *data mining* dapat mengidentifikasi masalah dan menganalisis suatu data sehingga dapat menentukan hasil keputusan dalam menentukan kelulusan. Data mining menggunakan teknik salah satunya klasifikasi. Klasifikasi adalah proses pengelompokan suatu himpunan objek dalam sebuah baris data dan dikelompokkan kedalam kelas yang berbeda-beda, Pohon keputusan akan sangat membantu dalam menentukan prediksi kelulusan siswa.

Dengan menggunakan pohon keputusan, informasi tentang faktor-faktor yang memengaruhi kelulusan dapat ditampilkan secara jelas dan terstruktur. Setiap cabang dalam pohon keputusan mewakili atribut dari data, sementara setiap daun menunjukkan hasil akhir, yaitu klasifikasi kelulusan siswa. Dengan cara ini, para pendidik dan pengelola pendidikan dapat dengan mudah memahami kriteria apa saja yang berkontribusi terhadap keberhasilan atau kegagalan siswa dalam ujian. (Awaluddin dkk., 2024)

Proses klasifikasi ini juga memungkinkan untuk mengidentifikasi siswa-siswa yang berisiko gagal, sehingga tindakan pencegahan atau bantuan dapat diberikan lebih awal. Misalnya, jika pohon keputusan menunjukkan bahwa nilai rapor, kehadiran, dan partisipasi aktif dalam kelas adalah atribut penting yang mempengaruhi kelulusan, maka perhatian lebih dapat diarahkan kepada siswa yang memiliki nilai rendah atau tingkat kehadiran yang tidak memuaskan.(Anwar & Ristasa, 2024) Selain itu, pemodelan menggunakan pohon keputusan juga memungkinkan untuk memperbarui dan meningkatkan algoritma berdasarkan data baru. Dengan semakin banyaknya data yang dikumpulkan dari tahun ke tahun, pohon keputusan dapat dioptimalkan untuk memberikan prediksi yang lebih akurat.

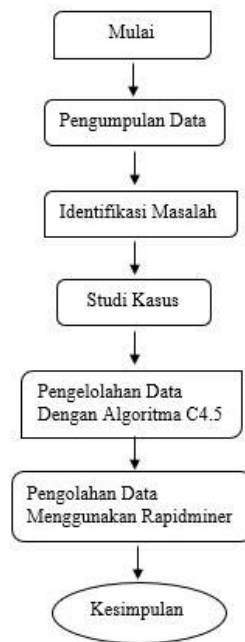
Hal ini tidak hanya meningkatkan proses pembelajaran siswa, tetapi juga membantu lembaga pendidikan dalam merumuskan kebijakan dan strategi yang lebih efektif untuk mendukung kelulusan siswa. Dengan demikian, penerapan teknik klasifikasi menggunakan pohon keputusan dalam data mining memberikan keuntungan yang

signifikan dalam memahami dan meningkatkan hasil pendidikan. Penelitian ini bertujuan untuk mengeksplorasi lebih dalam bagaimana algoritma C4.5 dapat menghasilkan model prediksi yang handal serta merekomendasikan langkah-langkah yang tepat untuk memaksimalkan peluang kelulusan siswa.

METODE

Untuk mendapatkan nilai perhitungan yang diperlukan sebagai hasil penting dari tahapan penelitian ini, telah disusun sebuah metode penelitian yang mencakup langkah-langkah seperti berikut ini:

Tahap Penelitian



Sumber : Hasil Penelitian (2024)
Gamabr 1. Metode Penelitian

- 1) Tahap Pra Lapangan
Pada tahap ini, proses penelitian yang dilakukan oleh peneliti, dimana penelitian akan mencari data dan mengidentifikasi suatu masalah yang ada pada data yang telah diperoleh.
- 2) Tahap Pekerjaan Lapangan
Pada tahap ini, yang juga dikenal sebagai tahap pengumpulan data, Dilakukan untuk memperoleh informasi yang dibutuhkan oleh peneliti dalam studi kasus untuk mencapai suatu tujuan peneliti. Data yang dikumpulkan oleh peneliti dari hasil observasi secara langsung di SDN Sukaresmi Kota Bogor.
- 3) Tahap Identifikasi Masalah
Dalam mengidentifikasi suatu masalah dengan menganalisis data untuk dilakukan penelitian. Hal ini identifikasi masalah terkasin dengan prediksi kelulusan siswa SDN Sukaresmi Kota Bogor.
- 4) Studi Pustaka
Studi Pustaka sebagai sumber referensi untuk mempelajari beberapa lieratur yang dapat digunakan dalam penelitian ini. dalam mencari studi pustaka merupakan

cara yang dilakukan untuk melengkapi pengetahuan dasar dan memahami teori-teori yang digunakan dalam penelitian ini.

- 5) Pengolahan data Dengan Alogaritma C4.5
pada prose pengolahan data dilakukan untuk menghasilkan suatu informasi yang akurat dan efisien guna untuk acuan dalam pengambilan suatu keputusan yang kemudian diolah menggunakan pohon keputusan Decision tree.
- 6) Kesimpulan
Dari data yang telah diambil dan diolah menjadi sampel data dalam pembahasan yang dilakukan maka dapat diperoleh suatu kesimpulan yang menjadi referensi untuk membuat suatu keputusan dan menjadi suatu referensi untuk penelitian selanjutnya. Pada tahapan ini pengolahan data yang diproses dan dihasilkan sebuah informasi yang dapat digunakan oleh sekolah tersebut.

Data Mining

Data Mining bukan sekedar terkumpul data saja tetapi mencakup analisis dan prediksi dari informasi yang ingin ditampilkan. "Data yang dikumpulkan disimpan dalam database kemudian diproses sehingga dapat dijadikan untuk pengambilan keputusan dalam melihat informasi yang akan digunakan" (Gaol dkk., 2021)

Algoritma C4.5

Dari beberapa algoritma yang dapat digunakan untuk membuat pohon keputusan (decision tree) yaitu diantaranya algoritma C4.5. Secara global algoritma C4.5 untuk menciptakan pohon keputusan sebagai berikut yaitu (Pangaribuan dkk., 2019) :

- 1) Menyiapkan data training
Data berasal dari data yang sudah dikumpulkan sebelumnya dan di kelompokkan kedalam kelas tertentu. (Zer dkk., 2022)
- 2) Menentukan akar dari pohon
Akar di pilih dari atribut dengan cara menghitung nilai gain dari semua atribut, dan nilai gain yang paling besar akan menjadi akar pertama. Sebelum menghitung nilai gain, sebaiknya hitung nilai entropy terlebih dahulu. (Siska Febriani & Heni Sulistiani, t.t.) Berikut merupakan rumus yang digunakan dalam menghitung nilai entropy :
Keterangan :
S = himpunan kasus
n = jumlah partisi S
pi = proporsi Si terhadap S
- 3) Hitung nilai gain dengan menggunakan rumus sebagai berikut :
$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$
 (Adriansa dkk., 2022)
Keterangan :
S = Himpunan Kasus A = Atribut
n = Jumlah Partisi Atribut A
| Si | = Jumlah Kasus pada partisi ke-i
| S | = Jumlah Kasus dalam S
- 4) Kemudian ulangi Langkah kedua sampai semua record terpartisi secara sempurna.
- 5) Proses partisi pohon keputusan akan berhenti jika :
 - a. Semua *record* pada simpul N mendapatkan kelas yang sama.
 - b. Tidak ada atribut di dalam record yang dipartisi lagi.
 - c. Tidak ada record didalam cabang yang kosong. (Ramadhan, 2022)

Pohon Keputusan (Decision Tree)

Pohon keputusan adalah metode klasifikasi yang menggunakan struktur pohon, di mana setiap node mewakili atribut dan cabangnya mewakili value dari atribut, sedangkan daun digunakan untuk mewakili kelas. Simpul atas pohon keputusan ini disebut root (Adinda, 2022)

RapidMiner

Data diolah menggunakan RapidMiner yang berguna sebagai validasi dan reabilitas data untuk mencari keakuratan data. "RapidMiner adalah pemrograman perangkat lunak ilmu data yang dikembangkan oleh Perusahaan Bersama-sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambangan teks, dan analisis prediktif" (Nurjana dkk., 2022)

HASIL DAN PEMBAHASAN

Dengan demikian hal inilah yang mendasari penulis dalam membuat jurnal mengenai Penerapan Data Mining Untuk Prediksi Kelulusan Siswa SDN Sukaresmi Kota Bogor Dengan Alogaritma C4.5

Data yang digunakan dalam membentuk pohon keputusan adalah data-data nilai siswa SDN Sukaresmi dengan kriteria-kriteria yang telah ditetapkan oleh SDN Sukaresmi dalam memprediksi kelulusan siswa SDN Sukaresmi. Selanjutnya akan dilakukan praproses untuk menghasilkan data kasus yang siap untuk dibentuk menjadi pohon keputusan. Pra proses data meliputi :

- 1) Seleksi data yaitu Langkah pertama adalah memilih subset data yang relevan dengan tujuan analisis. Seleksi ini bertujuan untuk mengidentifikasi variabel-variabel kunci yang memiliki pengaruh signifikan terhadap kelulusan siswa. Dengan memfokuskan pada data yang relevan, proses analisis akan menjadi lebih efisien dan hasil yang diperoleh akan lebih akurat. (Khaira dkk., 2020)
- 2) Pembersihan data yaitu Setelah melakukan seleksi, tahap berikutnya adalah membersihkan data dari kesalahan atau ketidaksesuaian, seperti data yang hilang, duplikasi, atau nilai yang tidak wajar. Proses pembersihan ini sangat penting untuk menghindari bias yang dapat muncul akibat data kotor atau tidak konsisten, sehingga kualitas model yang dibangun menjadi lebih tinggi. (Ahmad dkk., 2021)
- 3) Transformasi data yaitu Pada tahap ini, data yang telah dibersihkan akan ditransformasi ke dalam format yang lebih sesuai untuk analisis. Ini dapat mencakup pengkodean variabel kategorikal menjadi bentuk numerik, normalisasi atau standarisasi data, serta pembagian atribut berdasarkan skala yang tepat. Transformasi juga membantu dalam meningkatkan keterbacaan dan interpretasi dari pohon keputusan yang dihasilkan. (M.Kom & M.T, 2020)

No.	Nama	Nilai Harian	Nilai UTS	Nilai UAS	Perilaku
1	ABIMANYU AZKA WANAYASA	86	88	90	Baik
2	ADELIA NUR RAMADHANI	82	83	77	Baik
3	ALIKA NURZAINI	54	54	65	Cukup
4	ANNISA MAULIDA	63	60	70	Baik
5	DEVILA PUTRI SEPTIANI	60	63	60	Cukup
6	FLICKA AULIA HUSNA	84	85	89	Baik
7	ILHAM AL KADAFI	70	88	75	Baik
8	JULIANSYAH DHARMAWAN AL G.	91	91	83	Baik
9	MALIKHA ROZIYATUL LUTHFIAH	59	66	61	Cukup
10	M. RANDIKA ZHIO RAMADHAN	67	63	64	Baik
11	MUHAMAD RESTIAN PRAMANA	67	65	65	Baik
12	MUHAMMAD FAHMI	68	63	70	Baik

	APRIYANSYAH				
13	MUHAMMAD FAHRUL A'ZAM	70	69	68	Baik
14	MUHAMMAD KAHFI RAMADHAN	84	82	90	Baik

15	MUH. MULLAH ALY OMAR F. AL P.	68	66	66	Cukup
16	MUHAMMAD RIZKY PERDANA	70	71	76	Baik
17	MUHAMMAD TEGAR	79	93	80	Baik
18	NAYLA PUTRI RAHMANIA	83	89	87	Baik
19	NIA APRILLIYA	80	89	77	Baik
20	PUTRI NAYCHILLA RAMADANI	82	85	78	Baik
21	REZA ALAMSYAH	67	69	66	Cukup
22	SHINTA SUHARYANA	82	85	78	Baik
23	SITI NURI MAULIDA	66	64	65	Cukup
24	SOFIE ALICIA ZAHRA	90	89	93	Baik
25	VIRLY SATRIA PRATAMA	67	59	64	Cukup
26	YUSDINI LESTARI PANJAITAN	80	80	82	Baik
27	ZETISYA ALEYSYA RAHMAT	83	85	84	Baik
28	AMANDA PRAMUGIA PUTRI S.	68	70	69	Cukup
29	ANINDIA MAULIDA PUTRI	71	69	60	Cukup
30	BULAN PURNAMA RAHAYU SUGANDI	79	88	79	Baik
31	CLARA NATASYA	77	82	85	Baik
32	DANOE PUTRA PRATAMA	90	85	84	Baik
33	FADIL AZHAR KUSUMA ATMAJA	90	86	83	Baik
34	FENI PRIHATINI	84	87	79	Baik
35	HAURA HUMAIRA SALSABILA	70	70	75	Baik
36	KHEYLA NADZVILANA	75	70	78	Baik
37	M. KHAFIZ FAHYUDIN	76	84	82	Baik
38	MAULIDA TINNURI	92	87	80	Baik
39	MILIA RAMADHANI	94	94	96	Baik
40	MUHAMAD RIZKI MAULANA	52	55	54	Cukup
41	MUHAMMAD NAJIB BARIQ	58	54	58	Cukup
42	MUHAMMAD NIZAM SYABANA	85	75	75	Baik
43	MUHAMMAD RADITYA ANDRIANTO	79	79	78	Baik
44	MUH. RANDI RENALDI KHAIRI D.	69	68	64	Cukup
45	NABIL DHAIFULLAH	90	86	81	Baik
46	NAYELA CHAIRA NAJMIE	93	80	82	Baik
47	PUTRA SATRIA MAULANA	79	83	90	Baik
48	PUTRI QONITA KOSWARA	78	89	79	Baik
49	RAYHAN ALFA ATHARAUF R.	70	72	69	Baik
50	RIZQI DWI HIDAYATURRAMADHAN	65	61	66	Cukup
51	SALSABILA PUTRI ARYANTI	87	75	79	Baik
52	SYAFIRA NUR KAILA	91	75	87	Baik
53	SYIFA AULIA PUTRI	85	93	84	Baik
54	ZIDNI RAMADHAN	87	76	88	Baik
55	ZIFANA LETISYA	79	84	91	Baik

Atribut Sebagai Akar

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk mendapatkan nilai gain, terlebih dahulu harus mencari nilai entropy. Dengan menggunakan dua persamaan maka akan didapatkan nilai entropy dan gain yang digunakan sebagai akar dalam membuat pohon keputusan. Berikut ini merupakan rumus untuk mencari nilai entropy dan gain.

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

Gain adalah untuk Mengukur pengurangan entropy yang diperoleh dari pemisahan data berdasarkan atribut tertentu. (Hamzah dkk., 2024)

$$\text{Entropy}(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Entropy adalah untuk Mengukur ketidakpastian atau ketidakteraturan dalam data. (*Analisis Sentimen Ulasan Aplikasi Binar Pada Google Play Store Menggunakan Algoritma Naïve Bayes | INTECOMS: Journal of Information Technology and Computer Science*, t.t.) Dengan menerapkan kedua persamaan tersebut, nilai entropy dari seluruh dataset dapat dihitung terlebih dahulu untuk mengukur tingkat ketidakpastian. Selanjutnya, nilai gain untuk setiap atribut dapat dihitung untuk menilai seberapa efektif masing-masing atribut dalam mengurangi ketidakpastian tersebut saat digunakan untuk memisahkan data. Setelah memperoleh nilai gain untuk semua atribut, atribut dengan gain tertinggi akan dipilih sebagai akar dari pohon keputusan. Pemilihan akar yang tepat sangat penting, karena akar berfungsi sebagai titik awal dalam proses klasifikasi dan dapat menentukan arah dari keputusan yang diambil pada cabang-cabang berikutnya dalam pohon. (Zulkarnaen Noor Syarif, 2023) Proses ini terus berlanjut secara rekursif, mengulangi langkah yang sama untuk subset yang dihasilkan dari pemisahan sampai memenuhi kriteria pemberhentian, seperti mencapai kedalaman maksimum pohon atau ketika tidak ada atribut yang relevan tersisa. (Yanti & Eriana, 2024)

Dengan demikian, pendekatan berbasis entropy dan gain memberikan metode yang sistematis dalam menentukan struktur pohon keputusan yang optimal, menghasilkan model yang dapat memberikan prediksi yang lebih akurat terhadap kelulusan siswa berdasarkan atribut-atribut yang telah dipilih. (Nurhasanah Nurhasanah & Uden Saptoni, t.t.)

Berikut ini merupakan perhitungan nilai *entropy* :

1. *Entropy* Total

Entropy Total adalah menghitung nilai dari perilaku yaitu Baik dan Cukup. Dimana nilai *entropy* masing-masing adalah sebagai berikut :

$$\begin{aligned} \text{Entropy (Total)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-42/55 * \log_2 (42/55)) + (-13/55 * \log_2 (13/55)) \\ &= 0,788940657 \end{aligned}$$

2. *Entropy* Atribut Nilai Harian

Entropy Atribut Nilai harian terdiri dari 3 nilai yaitu nilai > 70, = 70, dan nilai < 70.

Dimana nilai *entropy* masing-masing adalah sebagai berikut :

$$\begin{aligned} \text{a. Entropy (> 70)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-33/34 * \log_2 (33/34)) + (-1/34 * \log_2 (1/34)) \\ &= 0,191433255 \\ \text{b. Entropy (= 70)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-5/5 * \log_2 (5/5)) + (-0/5 * \log_2 (0/5)) \\ &= 0 \\ \text{c. Entropy (< 70)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-4/16 * \log_2 (4/16)) + (-12/15 * \log_2 (12/15)) \\ &= 0,811278124 \end{aligned}$$

3. *Entropy* Atribut Nilai UTS

Entropy Atribut Nilai UTS terdiri dari 3 nilai yaitu nilai > 70, = 70, dan < 70. Dimana nilai *entropy* masing-masing adalah sebagai berikut.

$$\begin{aligned}
 \text{a. } \textit{Entropy} (> 70) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-35/35 \cdot \log_2 (35/35)) + (-0/35 \cdot \log_2 (0/35)) \\
 &= 0 \\
 \text{b. } \textit{Entropy} (= 70) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-2/3 \cdot \log_2 (2/3)) + (-1/3 \cdot \log_2 (1/3)) \\
 &= 0,918295834 \\
 \text{c. } \textit{Entropy} (< 70) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-5/17 \cdot \log_2 (5/17)) + (-12/17 \cdot \log_2 (12/17)) \\
 &= 0,873981048
 \end{aligned}$$

4. *Entropy* Atribut Nilai UAS

Entropy Atribut Nilai UAS terdiri dari 3 nilai yaitu nilai > 70, = 70, dan nilai < 70. Dimana nilai *entropy* masing-masing adalah sebagai berikut.

$$\begin{aligned}
 \text{a. } \textit{Entropy} (> 70) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-36/36 \cdot \log_2 (0/36)) + (-0/36 \cdot \log_2 (0/36)) \\
 &= 0 \\
 \text{b. } \textit{Entropy} (= 70) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-2/2 \cdot \log_2 (2/2)) + (-0/2 \cdot \log_2 (0/2)) \\
 &= 0 \\
 \text{c. } \textit{Entropy} (< 70) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-4/17 \cdot \log_2 (4/17)) + (-13/17 \cdot \log_2 (13/17)) \\
 &= 0,787126586
 \end{aligned}$$

Berikut ini merupakan perhitungan nilai *gain* :

1. *Gain* (Total, Nilai Harian)

$$\begin{aligned}
 \textit{Gain} (\text{Total, Nilai Harian}) &= \textit{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \textit{Entropy}(S_i) \\
 &= 0,788940657 - (35/55 * 0,315997133) \\
 &\quad + (-5/55 * 0) + (-15/55 * 0,836640742) \\
 &= 0,434591918
 \end{aligned}$$

2. *Gain* (Total, Nilai UTS)

$$\begin{aligned}
 \textit{Gain} (\text{Total, Nilai UTS}) &= \textit{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \textit{Entropy}(S_i) \\
 &= 0,788940657 - (37/55 * 0,179256067) \\
 &\quad + (-3/55 * 0,918295834) + (-15/55 * \\
 &\quad 0,836640742) \\
 &= 0,468712197
 \end{aligned}$$

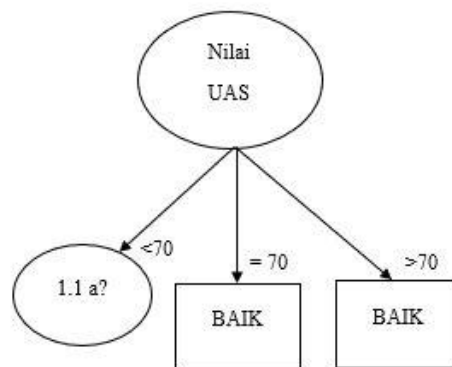
3. *Gain* (Total, Nilai UAS)

$$\begin{aligned}
 \textit{Gain} (\text{Total, Nilai UAS}) &= \textit{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \textit{Entropy}(S_i) \\
 &= 0,788940657 - (36/55 * 0) + (-2/5 * 0) + (-17/55 \\
 &\quad * \\
 &\quad 0,787126586) \\
 &= 0,545646985
 \end{aligned}$$

Setelah seluruh nilai *Entropy* dan *Gain* diperoleh, selanjutnya hasil dari perhitungan tersebut dimasukkan ke dalam Tabel sebagai berikut :

Node		Keterangan	Jumlah Kasus	Baik	Cukup	Entropy	Gain
1	Total		55	42	13	0,788940657	
	Nilai Harian						0,434591918
		> 70	34	33	1	0,191433255	
		= 70	5	5	0	0	
		< 70	16	4	12	0,811278124	
	Nilai UTS						0,468712197
		> 70	35	35	0	0	
		= 70	3	2	1	0,918295834	
		< 70	17	5	12	0,873981048	
	Nilai UAS						0,545646985
		> 70	36	36	0	0	
		= 70	2	2	0	0	
		< 70	17	4	13	0,787126586	

Pada perhitungan Tabel di atas dapat diketahui bahwa atribut dengan Gain tertinggi adalah Nilai UAS dengan nilai Gain 0,545646985 sehingga atribut Nilai UAS dapat menjadi Node Akar. Pada atribut Nilai UAS yang dijadikan node akar terdapat 3 nilai yaitu > 70, = 70 dan < 70. Dari Baik dan Sudah Mencapai dan untuk nilai (=70) mempunyai hasil = Baik dan Sudah Mencapai sehingga tidak perlu dilakukan perhitungan Kembali. Dan untuk nilai (< 70) mempunyai Hasil = Baik dan Hasil = Cukup sehingga untuk nilai < 70 harus dilakukan perhitungan lagi.



No.	Nama	Nilai Harian	Nilai UTS	Nilai UAS	Perilaku
1	MUH. RANDI RENALDI KHAIRI D.	69	68	64	Cukup
2	MUH. MULLAH ALY OMAR F. AL P.	68	66	66	Cukup
3	REZA ALAMSYAH	67	69	66	Cukup
4	AMANDA PRAMUGIA PUTRI S.	68	70	69	Cukup
5	ANINDIA MAULIDA PUTRI	71	69	60	Cukup
6	RIZQI DWI HIDAYATURRAMADHAN	65	61	66	Cukup
7	VIRLY SATRIA PRATAMA	67	59	64	Cukup
8	ALIKA NURZAINI	54	54	65	Cukup
9	MALIKHA ROZIYATUL LUTHFIAH	59	66	61	Cukup
10	MUHAMAD RIZKI MAULANA	52	55	54	Cukup
11	MUHAMMAD NAJIB BARIQ	58	54	58	Cukup
12	DEVILA PUTRI SEPTIANI	60	63	60	Cukup
13	SITI NURI MAULIDA	66	64	65	Cukup
14	RAYHAN ALFA ATHARAUF R.	70	72	69	Baik
15	MUHAMMAD FAHMI APRIYANSYAH	68	63	70	Baik
16	MUHAMMAD FAHRUL A'ZAM	70	69	68	Baik
17	MUHAMAD RESTIAN PRAMANA	67	65	65	Baik
18	M. RANDIKA ZHIO RAMADHAN	67	63	64	Baik
19	ANNISA MAULIDA	63	60	70	Baik

Cabang Untuk Setiap Nilai

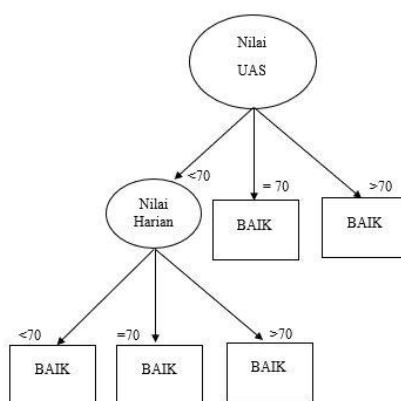
Untuk mencari *node* cabang berikutnya, maka harus dilakukan perhitungan Kembali seperti yang tertera diatas untuk mencari jumlah kasus dan nilai *entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut Nilai Harian, Nilai UTS, dan Nilai UAS yang dapat menjadi *node* cabang dari nilai atribut < 70. (Sungkar & Qurohman, 2021) Setelah itu, lakukan perhitungan nilai *gain* untuk tiap-tiap atribut. Hasil perhitungan ditunjukkan oleh tabel sebagai berikut :

Node		Keterangan	Jumlah Kasus	Baik	Cukup	Entropy	Gain
	Nilai UAS	< 70	17	4	13	0,787126586	
1.1							
	Nilai Harian						0,023570704
		> 70	1	0	1	0	
		70	2	2	0	0	
		< 70	16	4	12	0,811278124	
	Nilai UTS						-0,086854462

		> 70	1	1	0	0	
		70	1	0	1	0	
		< 70	17	5	12	0,873981048	

Dari hasil Tabel di atas dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah Nilai Harian, yaitu dengan nilai *Gain* 0,023570704. Dengan demikian atribut Nilai Harian dapat menjadi *node* cabang dari nilai atribut Baik. Pada atribut Nilai harian yang dijadikan *node* cabang terdapat nilai yaitu > 70, = 70, dan < 70. Dari nilai atribut tersebut, nilai > 70 mempunyai Hasil Baik = Sudah Mencapai dan nilai = 70 mempunyai Hasil Baik = Sudah Mencapai.

D. Pohon Keputusan



E. General Rules

Setelah pohon keputusan dibuat maka dibentuklah *rules* sebagai berikut :

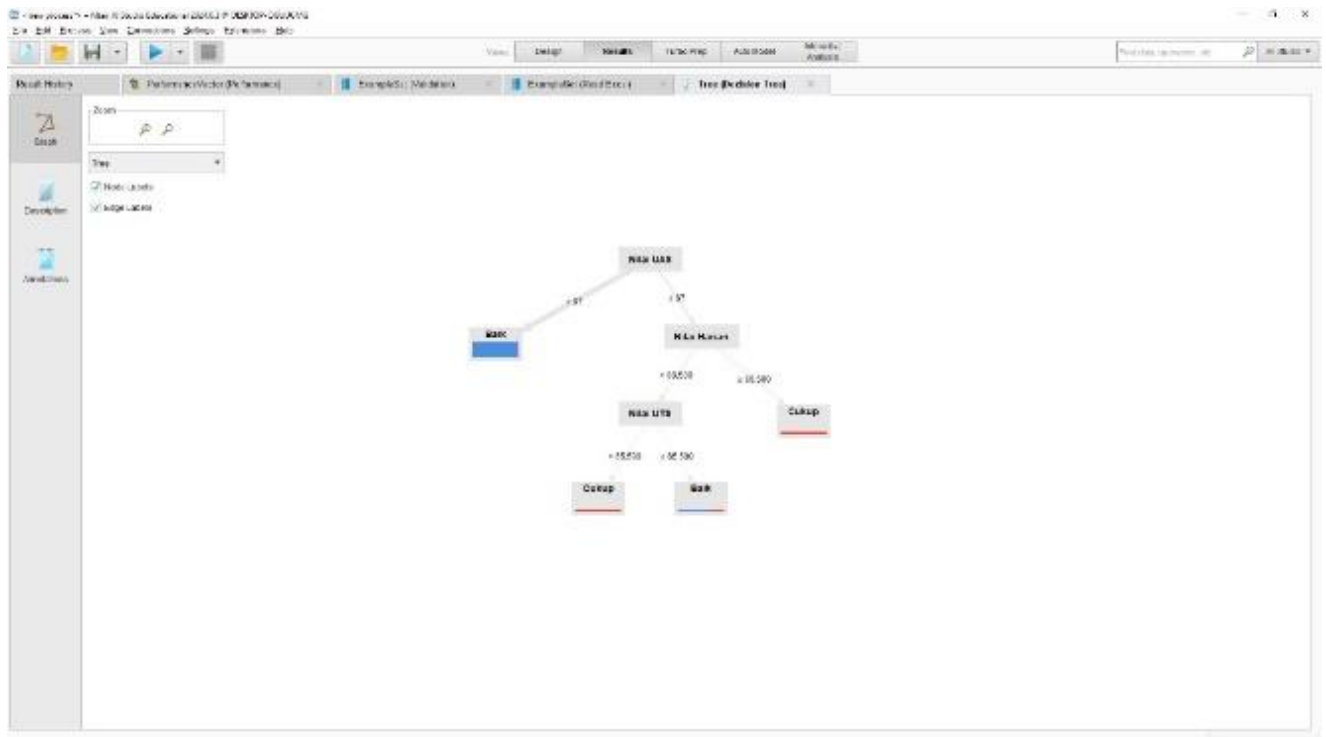
1. *If* Perilaku = Baik *and* Nilai Siswa = > 70 *then* di prediksi lulus.
2. *If* Perilaku = Baik *and* Nilai Siswa = = 70 *then* di prediksi lulus.
3. *If* Perilaku = Baik *and* Nilai Siswa = < 70 *then* di prediksi tidak lulus.
4. *If* Perilaku = Cukup *and* Nilai Siswa = < 70 *then* di prediksi tidak lulus.

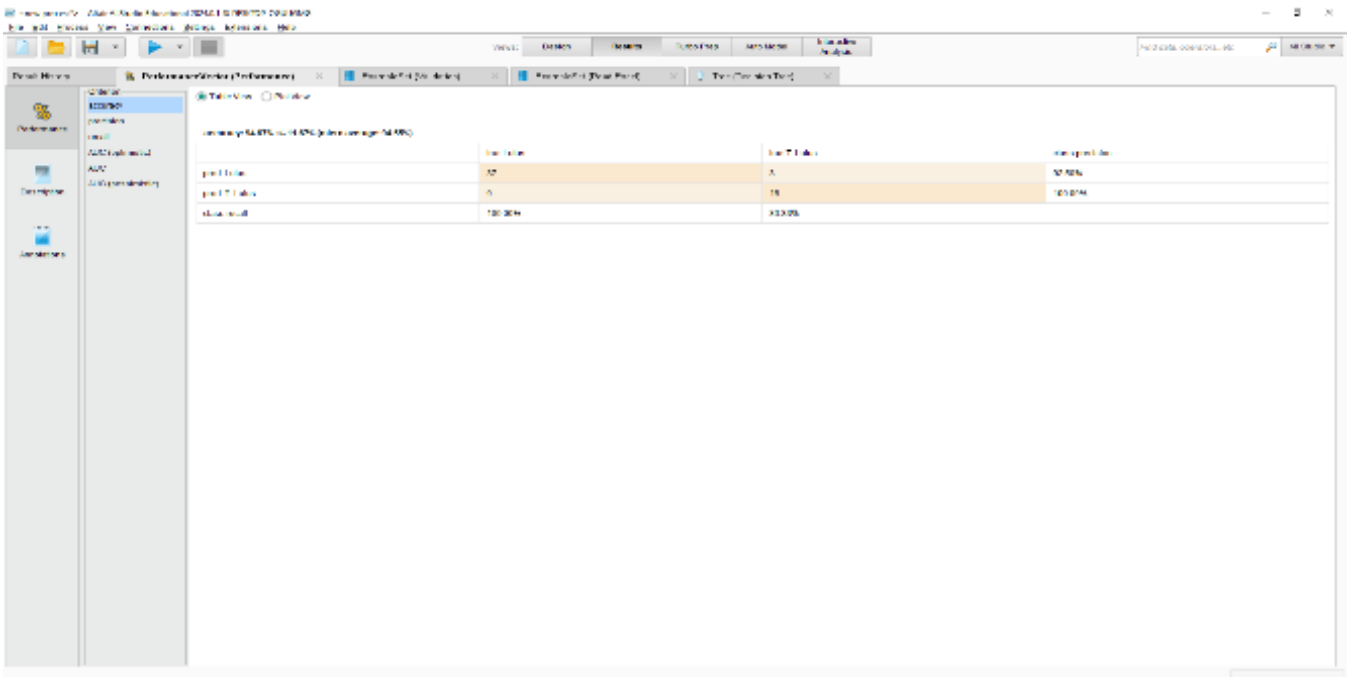
Berdasarkan perhitungan *node* dan pohon keputusan serta hasil *rules* yang dilakukan, maka diperoleh kesimpulan bahwa siswa yang diprediksi akan LULUS dan T. Lulus yaitu sebagai berikut :

No.	Nama	Nilai Harian	Nilai UTS	Nilai UAS	Perilaku	Keterangan
1	ABIMANYU AZKA WANAYASA	86	88	90	Baik	Lulus
2	ADELIA NUR RAMADHANI	82	83	77	Baik	Lulus
3	ALIKA NURZAINI	54	54	65	Cukup	T. Lulus
4	ANNISA MAULIDA	63	60	70	Baik	T. Lulus
5	DEVILA PUTRI SEPTIANI	60	63	60	Cukup	T. Lulus
6	FLICKA AULIA HUSNA	84	85	89	Baik	Lulus
7	ILHAM AL KADAFI	70	88	75	Baik	Lulus
8	JULIANSYAH DHARMAWAN AL G.	91	91	83	Baik	Lulus
9	MALIKHA ROZIYATUL LUTHFIAH	59	66	61	Cukup	T. Lulus

10	M. RANDIKA ZHIO RAMADHAN	67	63	64	Baik	T. Lulus
11	MUHAMAD RESTIAN PRAMANA	67	65	65	Baik	T. Lulus
12	MUHAMMAD FAHMI APRIYANSYAH	68	63	70	Baik	T. Lulus
13	MUHAMMAD FAHRUL A'ZAM	70	69	68	Baik	T. Lulus
14	MUHAMMAD KAHFI RAMADHAN	84	82	90	Baik	Lulus
15	MUH. MULLAH ALY OMAR F. AL P.	68	66	66	Cukup	T. Lulus
16	MUHAMMAD RIZKY PERDANA	70	71	76	Baik	Lulus
17	MUHAMMAD TEGAR	79	93	80	Baik	Lulus
18	NAYLA PUTRI RAHMANIA	83	89	87	Baik	Lulus
19	NIA APRILLIYA	80	89	77	Baik	Lulus
20	PUTRI NAYCHILLA RAMADANI	82	85	78	Baik	Lulus
21	REZA ALAMSYAH	67	69	66	Cukup	T. Lulus
22	SHINTA SUHARYANA	82	85	78	Baik	Lulus
23	SITI NURI MAULIDA	66	64	65	Cukup	T. Lulus
24	SOFIE ALICIA ZAHRA	90	89	93	Baik	Lulus
25	VIRLY SATRIA PRATAMA	67	59	64	Cukup	T. Lulus
26	YUSDINI LESTARI PANJAITAN	80	80	82	Baik	Lulus
27	ZETISYA ALEYSA RAHMAT	83	85	84	Baik	Lulus
28	AMANDA PRAMUGIA PUTRI S.	68	70	69	Cukup	T. Lulus
29	ANINDIA MAULIDA PUTRI	71	69	60	Cukup	T. Lulus
30	BULAN PURNAMA RAHAYU SUGANDI	79	88	79	Baik	Lulus
31	CLARA NATASYA	77	82	85	Baik	Lulus
32	DANOE PUTRA PRATAMA	90	85	84	Baik	Lulus
33	FADIL AZHAR KUSUMA ATMAJA	90	86	83	Baik	Lulus
34	FENI PRIHATINI	84	87	79	Baik	Lulus
35	HAURA HUMAIRA SALSABILA	70	70	75	Baik	Lulus
36	KHEYLA NADZVILANA	75	70	78	Baik	Lulus
37	M. KHAFIZ FAHYUDIN	76	84	82	Baik	Lulus
38	MAULIDA TINNURI	92	87	80	Baik	Lulus
39	MILIA RAMADHANI	94	94	96	Baik	Lulus
40	MUHAMAD RIZKI MAULANA	52	55	54	Cukup	T. Lulus
41	MUHAMMAD NAJIB BARIQ	58	54	58	Cukup	T. Lulus
42	MUHAMMAD NIZAM SYABANA	85	75	75	Baik	Lulus

43	MUHAMMAD RADITYA ANDRIANTO	79	79	78	Baik	Lulus
44	MUH. RANDI RENALDI KHAIRI D.	69	68	64	Cukup	T. Lulus
45	NABIL DHAIFULLAH	90	86	81	Baik	Lulus
46	NAYELA CHAIRA NAJMIE	93	80	82	Baik	Lulus
47	PUTRA SATRIA MAULANA	79	83	90	Baik	Lulus
48	PUTRI QONITA KOSWARA	78	89	79	Baik	Lulus
49	RAYHAN ALFA ATHARAUF R.	70	72	69	Baik	Lulus
50	RIZQI DWI HIDAYATURRAMADHAN	65	61	66	Cukup	T. Lulus
51	SALSABILA PUTRI ARYANTI	87	75	79	Baik	Lulus
52	SYAFIRA NUR KAILA	91	75	87	Baik	Lulus
53	SYIFA AULIA PUTRI	85	93	84	Baik	Lulus
54	ZIDNI RAMADHAN	87	76	88	Baik	Lulus
55	ZIFANA LETISYA	79	84	91	Baik	Lulus





PerformanceVector

PerformanceVector:

accuracy: 94.67% +/- 11.67% (micro average: 94.55%)

ConfusionMatrix:

True:	Lulus	T. Lulus
Lulus:	37	3
T. Lulus:	0	15

KESIMPULAN

Dari penjelasan di atas dapat diambil kesimpulan dalam penelitian ini, yaitu dari data yang diperoleh melalui riset secara langsung dengan pihak sekolah SDN Sukaresmi Kota Bogor tersedia 55 data dengan tipe data numerik dan nominal, data tersebut telah dioleh menggunakan metode klasifikasi dengan Algoritma C4.5 Pohon Keputusan (*Decision Tree*) dan diproses pada tools Rapidminer. Melalui *Pre-processing*, data training, permodelan algoritma C4.5 *Decision tree* terhadap data training, *apply model* terhadap data training dan dilakukan pengujian untuk menghasilkan data yang akurat terhadap data training. dari pengujian dalam penelitian ini dihasilkan data dinyatakan lulus semua dengan nilai 37 siswa dan dinyatakan tidak lulus dengan nilai 18 orang..

REFERENSI

- Adinda, P. R. (2022). Pohon Keputusan C4. 5 Algoritma Untuk Klasifikasi Program Bantuan Belajar. *Jurnal Portal Data*, 2(9), 1–13.
- Adriansa, M., Yulianti, L., & Elfianty, L. (2022). Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 115–121. <https://doi.org/10.54367/jtiust.v7i1.1983>
- Ahmad, F. Z., Arifandy, M. F. S., Caesarardhi, M. R., & Rakhmawati, N. A. (2021). Bagaimana Masyarakat Menyikapi Pembelajaran Tatap Muka: Analisis Komentar Masyarakat pada Media Sosial Youtube Menggunakan Algoritma Deep Learning Sekuensial dan LDA. *Jurnal Linguistik Komputasional*, 4(2), Article 2. <https://doi.org/10.26418/jlk.v4i2.57>
- Analisis Sentimen Ulasan Aplikasi Binar Pada Google Play Store Menggunakan Algoritma Naïve Bayes | INTECOMS: Journal of Information Technology and Computer Science.* (t.t.). Diambil 24 Oktober 2024, dari <https://journal.ipm2kpe.or.id/index.php/INTECOM/article/view/6117>
- Anwar, A., & Ristasa, R. (2024). Upaya Peningkatan Hasil Belajar Siswa Sekolah Menengah Pertama Melalui Penerapan Metode Inquiry pada Materi Klasifikasi MakhluK Hidup. *Journal of Science Biology and Online Learning*, 1(1), Article 1. <https://doi.org/10.33830/simbion.v1i1.9596>
- Awaluddin, M. R. N., Hamdani, H., Hartoyo, A., Bistari, B., & Siregar, N. (2024). Pengaruh kemampuan awal terhadap kemampuan berpikir kritis siswa MTs dalam pembelajaran matematika. *JPMI (Jurnal Pembelajaran Matematika Inovatif)*, 7(1), 205–216. <https://doi.org/10.22460/jpmi.v7i1.20189>
- Gaol, L. Y. L., Safii, M., & Suhendro, D. (2021). *Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi.* 2(2), 97–106.
- Hamzah, I., Sitorus, Z., & Khairul. (2024). Analisa Classification Decision Tree C45 dan Naïve Bayes Pada Indikasi Penyakit Diabetes Menggunakan Rapid Miner. *Jurnal*

- Nasional Teknologi Komputer*, 4(1), Article 1.
<https://doi.org/10.61306/jnastek.v4i1.126>
- Khaira, U., Syarif, N., & Hayati, I. (2020). *Prediksi Tingkat Fertilitas Pria Dengan Algoritma Pohon Keputusan Cart*. 5(1).
- Maqfiroh1, F., & Sri Mujiyono2. (2022). *PENERAPAN KLASIFIKASI ALGORITMA DATA MINING C4.5 UNTUK MEMPREDIKSI TINGKAT KELULUSAN SISWA DI LEMBAGA PELATIHAN KERJA SHINJU SEMARANG*. 20(1), 105–123.
- M.Kom, M. A., S. Si, & M.T, M. N., S. T. (2020). *Data Mining—Algoritma dan Implementasi*. Penerbit Andi.
- Nurhasanah Nurhasanah & Uden Saptoni. (t.t.). *Penerapan Algoritma Backtracking Dengan Bounding Function Dan Depth First Search Pada Permainan Boggle | Jurnal Teknik Informatika dan Elektro*. Diambil 24 Oktober 2024, dari <https://jurnal.ugp.ac.id/index.php/JURTIE/article/view/699>
- Nurjana, A., Perdana Windarto, A., Qurniawan, H., & Tunas Bangsa, S. (2022). Implementasi Data Mining Dalam Memprediksi Prestasi Siswa Dengan Algoritma C4.5. *SmartEDU*, 1(4), 171–180.
- Pangaribuan, J., Tedja, C., & Wibowo, S. (2019). Perbandingan Metode Algoritma C4.5 dan Extreme Learning Machine untuk Mendiagnosis Penyakit Jantng Korner. *Informatics Engineering Research and Technology*, 1(1), 1–7.
- Pascalina, D., Widhiastono, R., & Juliane, C. (2023). Pengukuran Kesiapan Transformasi Digital Smart City Menggunakan Aplikasi Rapid Miner. *Technomedia Journal*, 7(3 Februari), 293–302. <https://doi.org/10.33050/tmj.v7i3.1914>
- Putra, A. R. (2023). The Implementation of Data Mining Techniques for Predicting Student Study Period Using the C4.5 Algorithm: Penerapan Teknik Data Mining Terhadap Prediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5. *Indonesian Journal of Electrical Engineering and Renewable Energy (IJEERE)*, 3(2), Article 2. <https://doi.org/10.57152/ijeere.v3i2.986>
- Ramadhan, A. (2022). Sistem Pendukung Keputusan Evaluasi Problematika Pendampingan Pembelajaran Daring dengan Algoritma C4.5. *Jurnal Sistim Informasi Dan Teknologi*, 58–63. <https://doi.org/10.37034/jsisfotek.v4i2.124>
- Rofiani, R., Oktaviani, L., Vernanda, D., & Hendriawan, T. (2024). Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5. *Jurnal Tekno Kompak*, 18(1), Article 1. <https://doi.org/10.33365/jtk.v18i1.3525>
- Siska Febriani & Heni Sulistiani. (t.t.). *Analisis Data Hasil Diagnoga Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5 | Febriani | Jurnal Teknologi dan Sistem Informasi*. Diambil 24 Oktober 2024, dari <https://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/1373>
- Sungkar, M. S., & Qurohman, M. T. (2021). Penerapan Algoritma C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(3), Article 3. <https://doi.org/10.30865/mib.v5i3.3116>
- Tri Widiastuti, Koko Karsa, & Christina Juliane. (t.t.). *Evaluasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan Metode Klasifikasi Algoritma C4.5 | Technomedia Journal*. Diambil 24 Oktober 2024, dari <https://ijc.ilearning.co/index.php/TMJ/article/view/1932>
- Yanti, F., & Eriana, E. S. (2024). *Algoritma Sorting dan Searching*. Eureka Media Aksara. <https://repository.penerbiteitureka.com/publications/568014/>
- Zer, P. P. P. A. N. W. F. I. R. H., Hayadi, B. H., & Damanik, A. R. (2022). PENDEKATAN MACHINE LEARNING MENGGUNAKAN ALGORITMA C4.5 BERBASIS PSO DALAM ANALISA PEMAHAMAN PEMROGRAMAN WEBSITE. *Jurnal Informatika dan Teknik Elektro Terapan*, 10(3), Article 3. <https://doi.org/10.23960/jitet.v10i3.2700>

Zulkarnaen Noor Syarif. (2023). *Penerapan Information Gain Dan Algoritma K-Means Untuk Klasterisasi Kedisiplinan Pegawai Menggunakan Rapidminer | Syarif | TeknoIS: Jurnal Ilmiah Teknologi Informasi dan Sains.*
<http://teknois.unbin.ac.id/index.php/JBS/article/view/165>

Copyright Holder:

© Nurul Elvida, Adika May Sari (2024)

First Publication Right :

© Bulletin of Community Engagement

This article is under:

CC BY SA