



Pengaplikasian Algoritma C4.5 untuk Menganalisis Hubungan Kebiasaan Harian Siswa terhadap Prestasi Akademik

Muhammad Zikri Ansyaari^{1*}, Muhammad Yasin²

^{1,2}Teknik Informatika, Universitas Asahan, Indonesia

*Korespondensi Penulis: zikriansyar@gmail.com

Abstract. This study aims to analyze the relationship between students' daily habits and their academic performance using the C4.5 algorithm. The daily habit variables examined include study duration and intensity, sleep quality and patterns, frequency and type of gadget use, attendance consistency, and students' engagement in learning activities both inside and outside the classroom. The data were collected through questionnaires and combined with students' academic grades as indicators of performance. The C4.5 algorithm was employed to construct a decision tree model capable of identifying the daily habit attributes that have the most significant influence on academic achievement. The findings reveal that study intensity, sleep quality, and attendance consistency serve as dominant factors in predicting students' performance levels. Furthermore, the resulting decision tree model demonstrates a relatively high accuracy level, making it a useful tool for evaluation, student development planning, and decision-making processes by educators. These results confirm that the application of the C4.5 algorithm is effective in uncovering the relationship patterns between daily habits and academic achievement and has the potential to serve as a reference for efforts to improve students' learning quality.

Keywords: Academic Achievement; C4.5 Algorithm; Decision Tree; Learning Behavior Analysis; Students' Daily Habits

Abstrak. Penelitian ini bertujuan untuk menganalisis hubungan antara berbagai kebiasaan harian siswa dengan prestasi akademik mereka melalui pemanfaatan algoritma C4.5. Variabel kebiasaan harian yang dikaji mencakup durasi dan intensitas belajar, kualitas dan pola tidur, frekuensi serta jenis penggunaan gawai, tingkat kehadiran, serta keterlibatan siswa dalam aktivitas pembelajaran baik di dalam maupun di luar kelas. Data penelitian diperoleh melalui pengisian kuesioner yang kemudian dikombinasikan dengan nilai akademik siswa sebagai acuan penilaian prestasi. Algoritma C4.5 digunakan untuk membangun model pohon keputusan yang mampu mengidentifikasi atribut-atribut kebiasaan harian yang memiliki pengaruh paling signifikan terhadap capaian akademik. Hasil penelitian menunjukkan bahwa intensitas belajar, kualitas tidur, dan konsistensi kehadiran merupakan faktor dominan dalam memprediksi tingkat prestasi siswa. Selain itu, model pohon keputusan yang dihasilkan memiliki tingkat akurasi yang cukup tinggi, sehingga dapat dimanfaatkan sebagai alat bantu dalam proses evaluasi, perencanaan pembinaan, dan pengambilan keputusan oleh pendidik. Temuan ini menegaskan bahwa algoritma C4.5 efektif dalam mengungkap pola hubungan kebiasaan harian dan prestasi akademik, serta berpotensi menjadi rujukan dalam upaya peningkatan mutu belajar siswa.

Kata kunci: Algoritma C4.5; Analisis Perilaku Belajar; Kebiasaan Harian Siswa; Pohon Keputusan; Prestasi Akademik

1. LATAR BELAKANG

Prestasi akademik merupakan indikator penting dalam menilai keberhasilan proses pembelajaran siswa di lingkungan sekolah. Pencapaian akademik yang baik tidak hanya mencerminkan tingkat penguasaan siswa terhadap materi pelajaran tetapi juga menjadi tolok ukur dalam mengevaluasi mutu pendidikan. Berbagai faktor dapat memengaruhi prestasi akademik, salah satunya adalah kebiasaan harian siswa.

Kebiasaan seperti durasi belajar di rumah, waktu tidur, intensitas penggunaan gadget serta pola aktivitas harian lainnya diyakini memiliki pengaruh langsung terhadap tingkat konsentrasi, daya serap materi dan performa akademik siswa. Namun di lapangan sering

dijumpai bahwa banyak siswa belum mampu mengelola kebiasaan mereka secara optimal. Misalnya sebagian besar siswa lebih banyak menghabiskan waktu untuk bermain media sosial atau gim dibandingkan belajar, memiliki jam tidur yang tidak teratur serta kurang disiplin dalam mengatur waktu belajar. Kondisi ini secara perlahan berdampak pada penurunan prestasi akademik yang tercermin dari hasil nilai ujian dan penilaian harian.

Di lingkungan SMA Tamansiswa Sukadarmi khususnya pada jenjang kelas X hingga XII, guru sering kali melakukan analisis terhadap hubungan antara kebiasaan harian dan prestasi akademik siswa secara manual melalui observasi dan pencatatan nilai. Namun metode tersebut memiliki beberapa keterbatasan antara lain memerlukan waktu dan tenaga yang cukup besar, serta rentan terhadap subjektivitas penilai, sehingga hasilnya kurang konsisten dan sulit dijadikan dasar dalam pengambilan keputusan strategis.

Seiring berkembangnya teknologi informasi dan ilmu data, muncul kebutuhan akan sistem yang mampu mengolah data kebiasaan siswa secara otomatis serta menghasilkan analisis yang akurat dan objektif. Penerapan algoritma C4.5 sebagai salah satu metode dalam *data mining* dapat menjadi solusi yang tepat. Algoritma C4.5 merupakan metode *decision tree* (pohon keputusan) yang bekerja dengan membentuk aturan klasifikasi berdasarkan atribut-atribut yang terdapat dalam dataset. Dengan menggunakan algoritma ini, pola hubungan antara kebiasaan harian siswa dan prestasi akademik dapat digambarkan secara sistematis, objektif, serta mudah dipahami.

Penelitian terdahulu oleh (Firdaus et al., 2023) yang berjudul Penerapan Metode C4.5 untuk Klasifikasi Potensi Akademik Siswa menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan potensi akademik siswa berdasarkan nilai-nilai mata pelajaran dengan akurasi mencapai 93,33%. Namun, penelitian ini masih subjektivitas penilai, sehingga hasilnya kurang konsisten dan sulit dijadikan dasar dalam pengambilan keputusan strategis.

Seiring berkembangnya teknologi informasi dan ilmu data, muncul kebutuhan akan sistem yang mampu mengolah data kebiasaan siswa secara otomatis serta menghasilkan analisis yang akurat dan objektif. Penerapan algoritma C4.5 sebagai salah satu metode dalam *data mining* dapat menjadi solusi yang tepat. Algoritma C4.5 merupakan metode *decision tree* (pohon keputusan) yang bekerja dengan membentuk aturan klasifikasi berdasarkan atribut-atribut yang terdapat dalam dataset. Dengan menggunakan algoritma ini, pola hubungan antara kebiasaan harian siswa dan prestasi akademik dapat digambarkan secara sistematis, objektif, serta mudah dipahami.

Penelitian terdahulu oleh (Firdaus et al., 2023) yang berjudul Penerapan Metode C4.5 untuk Klasifikasi Potensi Akademik Siswa menunjukkan bahwa algoritma C4.5 mampu

mengklasifikasikan potensi akademik siswa berdasarkan nilai-nilai mata pelajaran dengan akurasi mencapai 93,33%. Namun, penelitian ini masih terbatas pada data akademik, tanpa mempertimbangkan aspek kebiasaan harian yang lebih kompleks dan dinamis. terbatas pada data akademik, tanpa mempertimbangkan aspek kebiasaan harian yang lebih kompleks dan dinamis.

2. KAJIAN TEORITIS

Data Mining

Data mining adalah proses penggalian dan analisis data dalam jumlah besar untuk menemukan pola, hubungan, tren, atau informasi tersembunyi yang berguna dalam pengambilan keputusan. Proses ini melibatkan penerapan teknik statistik, kecerdasan buatan, dan *machine learning* untuk mengekstraksi pengetahuan dari data mentah yang tersimpan dalam basis data atau sumber informasi lainnya. Tujuan utama *data mining* adalah mengubah data yang kompleks dan beragam menjadi informasi yang bermakna serta dapat digunakan untuk memprediksi perilaku, mengidentifikasi peluang dan mendukung strategi bisnis maupun penelitian ilmiah.

Data mining merupakan kegiatan menelusuri dan menemukan pola atau informasi yang relevan dari kumpulan data yang telah dipilih, dengan memanfaatkan metode atau teknik tertentu. Data yang diolah akan menghasilkan sebuah pola yang kemudian akan membantu pengambilan keputusan. *Data mining* adalah proses yang mencakup pengumpulan serta pemanfaatan data historis untuk mengidentifikasi keteraturan maupun pola hubungan di dalam himpunan data yang berukuran besar (Jauhar et al., 2021).

Machine Learning

Machine learning adalah cabang dari kecerdasan buatan (*artificial intelligence*) yang berfokus pada pengembangan algoritma dan model yang memungkinkan komputer untuk belajar dari data tanpa harus diprogram secara eksplisit. Sistem *machine learning* dapat mengenali pola, membuat prediksi, dan mengambil keputusan secara otomatis berdasarkan pengalaman sebelumnya.

Menurut (Nurhalizah et al., 2024), *machine learning* adalah komponen utama dalam bidang kecerdasan buatan yang digunakan untuk menyelesaikan berbagai permasalahan. Ini merupakan implementasi dari kecerdasan buatan yang fokus pada pengembangan sistem yang dapat belajar sendiri tanpa perlu diprogram secara berulang. *Machine Learning* adalah bidang penelitian ilmiah yang mempelajari algoritma dan model statistik yang digunakan oleh sistem

komputer untuk menyelesaikan tugas-tugas tertentu tanpa instruksi eksplisit, melainkan dengan memanfaatkan pola dan inferensi sebagai penggantinya

Tujuan utama *machine learning* adalah memungkinkan komputer untuk meningkatkan kinerjanya seiring bertambahnya data dan pengalaman, sehingga dapat digunakan dalam berbagai aplikasi seperti klasifikasi, prediksi, pengenalan gambar, dan analisis perilaku.

Algoritma C4.5

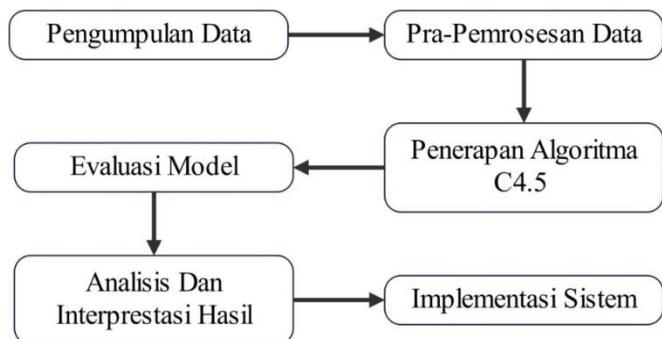
Algoritma C4.5 adalah salah satu algoritma pohon keputusan (*decision tree*) yang digunakan untuk klasifikasi dan prediksi dalam *data mining* dan *machine learning*. Algoritma ini merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser 3*) yang diperkenalkan oleh Ross Quinlan.

Menurut (Bachtiar et al., 2023), Algoritma C4.5 adalah program yang memberi kontribusi satu set data berlabel dan menghasilkan pohon keputusan sebagai keluaran. Pohon keputusan tindak lanjut ini kemudian diverifikasi terhadap data uji berlabel yang tidak terlihat untuk menghitung generalisasinya.

Algoritma C4.5 yaitu sebuah algoritma yang digunakan untuk membangun *decision tree* (pengambilan keputusan). Algoritma C4.5 adalah salah satu algoritma induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel training, label training dan atribut (Tukino, 2019).

3. METODE PENELITIAN

Beberapa langkah kerangka penelitian yang dibuat untuk rancangan sistem pada analisis hubungan kebiasaan harian siswa terhadap prestasi akademik menggunakan algoritma C4.5.



Gambar 1. Kerangka Kerja Penelitian

Uraian Kerangka Kerja

Berdasarkan kerangka kerja penelitian yang telah digambarkan di atas, maka pembahasan setiap tahapan dalam penelitian ini dapat diuraikan sebagai berikut:

Pengumpulan Data

Tahap awal penelitian adalah mengumpulkan data yang relevan dengan fokus penelitian. Data utama mencakup kebiasaan harian siswa misalnya durasi belajar, jam tidur, intensitas penggunaan media digital, waktu bermain, dan aktivitas sehari-hari lainnya, serta data prestasi akademik siswa SMA Tamansiswa Sukadamai seperti nilai rapor atau hasil ujian. Data diperoleh melalui penyebaran kuesioner dan dokumentasi dari sistem informasi sekolah.

Pra-pemrosesan Data

Langkah selanjutnya adalah pra-pemrosesan data (*data preprocessing*). Tahapan ini dilakukan untuk memastikan kualitas data sebelum dianalisis lebih lanjut. Proses yang dilakukan meliputi pembersihan data dari nilai kosong atau duplikat, normalisasi untuk menyeragamkan skala atribut, serta seleksi fitur guna memilih atribut yang paling relevan terhadap tujuan analisis. Hasil akhir dari tahap ini adalah data yang bersih, terstruktur, dan siap digunakan pada proses penerapan algoritma.

Penerapan Algoritma C4.5

Pada tahap ini dilakukan proses pembangunan model analisis menggunakan algoritma C4.5. Algoritma ini bekerja dengan menghitung nilai *Entropy* dan *Information Gain* dari setiap atribut untuk membentuk pohon keputusan. Atribut dengan nilai *Gain* tertinggi akan dipilih sebagai simpul utama dan proses ini berlanjut hingga terbentuk struktur pohon keputusan yang optimal. Model yang dihasilkan akan menjadi dasar untuk melakukan analisis terhadap data.

Evaluasi Model

Setelah model terbentuk dilakukan evaluasi terhadap performanya. Evaluasi bertujuan untuk mengukur sejauh mana model mampu melakukan analisis terhadap data secara benar. Ukuran kinerja yang digunakan dapat berupa akurasi, presisi, dan recall. Dari tahap ini dapat diketahui seberapa baik algoritma C4.5 bekerja terhadap dataset yang digunakan serta apakah model tersebut layak untuk diimplementasikan lebih lanjut.

Analisis Dan Interpretasi Hasil

Hasil dari evaluasi model kemudian dianalisis untuk mengetahui pola atau hubungan yang terbentuk dari data. Analisis dilakukan untuk memahami faktor-faktor yang paling memengaruhi hasil analisis serta bagaimana hasil tersebut dapat dimanfaatkan dalam konteks penelitian. Interpretasi hasil juga berfungsi sebagai dasar dalam memberikan rekomendasi atau pengambilan keputusan yang berbasis data.

Implementasi Sistem

Tahap terakhir adalah implementasi sistem yaitu penerapan model sistem ke dalam bentuk aplikasi atau sistem berbasis web. Implementasi ini bertujuan agar hasil penelitian dapat dimanfaatkan secara praktis oleh pengguna. Sistem yang dibangun memungkinkan pengguna untuk memasukkan data baru dan memperoleh hasil analisis secara otomatis. Dengan demikian penelitian ini tidak hanya menghasilkan model analisis tetapi juga sistem yang aplikatif dan bermanfaat.

Metode Penelitian

Dalam penelitian ini, metode yang digunakan meliputi studi pustaka dan penerapan algoritma klasifikasi C4.5 untuk menganalisis hubungan antara kebiasaan harian siswa dengan prestasi akademik mereka. Studi pustaka dilakukan untuk memperoleh pemahaman teori terkait kebiasaan harian siswa, indikator prestasi akademik, serta konsep dan implementasi algoritma C4.5. Algoritma C4.5 digunakan sebagai pendekatan utama dalam proses yang bertujuan membentuk pohon keputusan berdasarkan atribut kebiasaan siswa, sehingga dapat diperoleh pola yang menjelaskan pengaruh kebiasaan tersebut terhadap tingkat prestasi akademik.

4. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil yang diperoleh dari proses penelitian mulai dari analisis data kebiasaan harian siswa, perancangan sistem klasifikasi prestasi akademik hingga implementasi dan pengujian model berbasis web. Dataset yang digunakan dikumpulkan melalui penyebaran kuesioner kepada siswa SMA Tamansiswa Sukadarmi dengan atribut utama seperti durasi belajar, waktu tidur, intensitas penggunaan gadget dan nilai akademik siswa yang berhubungan dengan proses belajar.

Model klasifikasi dalam sistem ini dibangun menggunakan metode C4.5 yang bekerja dengan membentuk pohon keputusan berdasarkan dari setiap atribut. Metode ini mampu memisahkan data siswa berdasarkan pola kebiasaan harian sehingga setiap kategori prestasi akademik dapat diidentifikasi secara lebih akurat. Hasil pengujian menunjukkan bahwa sistem mampu mengklasifikasikan tingkat prestasi akademik siswa secara otomatis dengan tingkat akurasi yang baik.

Sistem yang dibangun juga mempermudah pengguna dalam menganalisis kebiasaan harian siswa tanpa perlu melakukan perhitungan atau pengamatan manual. Dengan adanya model klasifikasi pihak sekolah dapat lebih cepat memahami faktor-faktor yang memengaruhi prestasi serta membuat langkah pembinaan yang lebih tepat sasaran.

Tabel 1. Data Kebiasaan Siswa

No	Nama	Durasi Belajar	Durasi Tidur	Durasi Gadget	Nilai
1	Dian Marsayanti	1 - 3 Jam	6 - 8 Jam	2 - 5 Jam	> 81
2	Ela Saputri	< 1 Jam	> 8 Jam	< 2 Jam	61 - 80
3	Sindi Rizky	1 - 3 Jam	> 8 Jam	2 - 5 Jam	> 81
4	Meuthia Hafid	1 - 3 Jam	> 8 Jam	2 - 5 Jam	> 81
5	Aini Hasibuan	< 1 Jam	> 8 Jam	> 5 Jam	61 - 80
6	Felix Fernando	< 1 Jam	6 - 8 Jam	> 5 Jam	> 81
7	Nadin Wulan Cintya	> 3 Jam	> 8 Jam	> 5 Jam	> 81
8	Dinda Aurelia	1 - 3 Jam	6 - 8 Jam	2 - 5 Jam	> 81
9	Rara Anjani	1 - 3 Jam	< 6 Jam	2 - 5 Jam	> 81
10	Icha Adetia Sinaga	< 1 Jam	6 - 8 Jam	2 - 5 Jam	> 81
11	Nurmala Sari	1 - 3 Jam	< 6 Jam	< 2 Jam	> 81
12	Aziz Saputra	< 1 Jam	< 6 Jam	> 5 Jam	61 - 80
13	Nur Fadillah	> 3 Jam	< 6 Jam	2 - 5 Jam	61 - 80
14	Putri Natalia	1 - 3 Jam	< 6 Jam	< 2 Jam	> 81
15	M. Ikbal Al Af Dani	< 1 Jam	> 8 Jam	2 - 5 Jam	61 - 80
16	Syalsa Abelia	< 1 Jam	< 6 Jam	2 - 5 Jam	61 - 80
17	Nurhazizah	< 1 Jam	> 8 Jam	< 2 Jam	> 81
18	Nurhazizah	< 1 Jam	< 6 Jam	< 2 Jam	> 81
19	Reza Zikry Ramadhan	< 1 Jam	< 6 Jam	> 5 Jam	> 81
20	Aditya Anggara	> 3 Jam	> 8 Jam	> 5 Jam	> 81
21	Asti Fitri Ainiyah	1 - 3 Jam	6 - 8 Jam	> 5 Jam	> 81
22	Bina Wahyunisyah	> 3 Jam	6 - 8 Jam	> 5 Jam	> 81
23	Citra Nareen Kinayati	1 - 3 Jam	6 - 8 Jam	2 - 5 Jam	> 81
24	Daya Anggraini	< 1 Jam	> 8 Jam	< 2 Jam	61 - 80
25	Dwi Utami	1 - 3 Jam	> 8 Jam	2 - 5 Jam	> 81
26	Fadillah Sitorus	< 1 Jam	> 8 Jam	2 - 5 Jam	> 81
27	Fani Lestari	1 - 3 Jam	> 8 Jam	> 6 Jam	61 - 80
28	Farin Lusyahbana	1 - 3 Jam	> 8 Jam	> 6 Jam	> 81
29	Gadis Lestari	< 1 Jam	> 8 Jam	> 6 Jam	> 81
30	Hodijjah Sinaga Intan	< 1 Jam > 3 Jam	> 8 Jam < 6 Jam	2 - 5 Jam 2 - 5 Jam	> 81 > 81
31	Zulhasmendra				
32	Khoidijah Sinaga	1 - 3 Jam	6 - 8 Jam	2 - 5 Jam	> 81
33	Muhammad Akbar Rahmadan	1 - 3 Jam	< 6 Jam	< 2 Jam	> 81
34	Muhammad Azri	< 1 Jam	< 6 Jam	> 6 Jam	61 - 80
35	Naufal Rizki	1 - 3 Jam	< 6 Jam	2 - 5 Jam	61 - 80
36	Naufal Rizki Amar	< 1 Jam	< 6 Jam	3 - 5 Jam	> 81
37	Palentin Nithya	> 3 Jam	> 8 Jam	< 2 Jam	61 - 80
38	Reza Nigraputra	1 - 3 Jam	< 6 Jam	2 - 5 Jam	> 82

39	Restu Cahyo Purnomo	< 1 Jam	> 8 Jam	2 - 5 Jam	62 - 80
40	Ryan Siahaan	< 1 Jam	< 6 Jam	> 5 Jam	> 81
41	Saskia Maulida Ramadhani Saragih	< 1 Jam	< 6 Jam	> 5 Jam	> 81
42	Tegar Fathir Ramadhan	< 1 Jam	> 8 Jam	> 5 Jam	62 - 80
43	Velisa Afheyna Manurung	< 1 Jam	6 - 8 Jam	2 - 5 Jam	> 81
44	Putri Mheila	> 3 Jam	6 - 8 Jam	2 - 5 Jam	> 81
45	Ade Amanda	1 - 3 Jam	6 - 8 Jam	2 - 5 Jam	> 81
46	Aldo Ramadhan Rakasiwi	> 3 Jam	> 8 Jam	< 2 Jam	> 81
47	Alfahri Ramadhan	1 - 3 Jam	> 8 Jam	> 5 Jam	> 81
48	Alisha Putri	< 1 Jam	> 8 Jam	3 - 5 Jam	> 81
49	Awang Abryan	1 - 3 Jam	> 8 Jam	< 2 Jam	61 - 80
50	Azri Ade Syabana	< 1 Jam	> 8 Jam	2 - 5 Jam	61 - 80
51	Bunga Pebriola Tampubolon	1 - 3 Jam	> 8 Jam	2 - 5 Jam	> 81
52	Davinka Ramadhan	1 - 3 Jam	> 8 Jam	< 2 Jam	61 - 80
53	Dewi Yuliani	< 1 Jam	< 6 Jam	< 2 Jam	61 - 80
54	Fathan Tri Alfarizi	< 1 Jam	> 8 Jam	> 5 Jam	> 81
55	Fita Al Sifah	> 3 Jam	< 6 Jam	> 5 Jam	> 81
56	Fitri Anggun Lestari	1 - 3 Jam	< 6 Jam	> 5 Jam	> 81
57	Irvan Alnur Ridho	1 - 3 Jam	< 6 Jam	> 5 Jam	> 81
58	Jessica Anastasya Sihombing	< 1 Jam	< 6 Jam	3 - 5 Jam	> 81
59	Josua Gofinda Simatupang	1 - 3 Jam	> 8 Jam	< 2 Jam	> 81
60	Klaira Adelia Putri	< 1 Jam	< 6 Jam	2 - 5 Jam	> 81
61	Mitra Rizki Pranata	> 3 Jam	> 8 Jam	2 - 5 Jam	62 - 80
62	Muhammad Raihan Fahrezi	1 - 3 Jam	< 5 Jam	> 6 Jam	> 81
63	Naila Dzahabiyyah	< 1 Jam	< 6 Jam	> 6 Jam	> 81
64	Nazril Utama	< 1 Jam	> 8 Jam	> 6 Jam	62 - 80
65	Nisa Kamila	< 1 Jam	> 8 Jam	2 - 5 Jam	> 81
66	Nita Rosalina	< 1 Jam	> 8 Jam	2 - 5 Jam	> 81
67	Nur Nabila	< 1 Jam	> 8 Jam	2 - 5 Jam	> 81

68	Rafi Sinaga	> 3 Jam	> 8 Jam	< 2 Jam	> 81
69	Rendi Maulana	1 - 3 Jam	> 8 Jam	> 6 Jam	> 81
70	Revan Alhadi	> 3 Jam	> 8 Jam	3 - 5 Jam	> 81
71	Revan Juli Aditya	1 - 3 Jam	> 8 Jam	4 - 5 Jam	60 - 80
72	Rizky Asri Arrasyah	< 1 Jam	> 8 Jam	< 2 Jam	60 - 80
73	Rizky Butar- Butar	1 - 3 Jam	> 8 Jam	2 - 5 Jam	> 81
74	Seriyan Utama	< 1 Jam	> 8 Jam	2 - 5 Jam	60 - 80
75	Setiyo Pramana Putra	1 - 3 Jam	> 8 Jam	> 5 Jam	> 83
76	Sifa Aqila Nazua	1 - 3 Jam	> 8 Jam	> 5 Jam	60 - 80
77	Suzuya Hadista	< 1 Jam	< 6 Jam	> 5 Jam	> 81
78	Tria Selvia	< 1 Jam	< 6 Jam	2 - 5 Jam	> 81
79	Wildan Ramadhan	> 3 Jam	< 6 Jam	2 - 5 Jam	60 - 80
80	Yogi Pratama	1 - 3 Jam	< 6 Jam	2 - 5 Jam	> 81

Pembersihan Data

Tahap pembersihan data merupakan proses awal sebelum data dapat digunakan dalam pelatihan dan pengujian model klasifikasi. Langkah pertama yang dilakukan adalah pemeriksaan struktur dataset untuk memahami bentuk dan jenis data yang digunakan. Dataset yang digunakan memiliki atribut utama yaitu durasi belajar, durasi tidur, durasi penggunaan gadget serta nilai prestasi akademik siswa. Pemeriksaan dilakukan untuk memastikan setiap atribut berada dalam format yang benar, tidak ada kolom yang tertukar dan seluruh nilai dapat diproses dengan baik oleh sistem.

Proses pembersihan dilakukan untuk mengatasi ketidakkonsistenan penulisan maupun kesalahan input yang ditemukan pada dataset. Pada atribut durasi belajar dan durasi tidur misalnya terdapat variasi penulisan seperti “1 - 3 Jam”, “1-3 Jam” dan “1 – 3 Jam”, yang kemudian distandardkan menjadi format yang seragam. Pada atribut penggunaan gadget juga diperbaiki penulisan seperti “2 - 5 Jam” atau “> 5 Jam” agar tidak dianggap sebagai kategori yang berbeda.

Pada atribut prestasi akademik ditemukan adanya karakter khusus seperti “60 80” atau “>82” akibat kesalahan. Nilai-nilai tersebut diperbaiki menjadi “60 – 80” dan “>81” agar sesuai dengan label yang semestinya.

Atribut kategorikal seperti durasi belajar, durasi tidur dan durasi penggunaan gadget kemudian diubah menjadi nilai numerik melalui proses encoding untuk memudahkan algoritma C4.5 dalam membaca dan mengolah data. Contohnya durasi belajar diubah menjadi “< 1 Jam” = 0, “1 – 3 Jam” = 1 dan “> 3 Jam” = 2. Pada durasi tidur digunakan pengkodean seperti “<

Jam” = 0, “6 – 8 Jam” = 1 dan “> 8 Jam” = 2. Sedangkan penggunaan gadget dikodekan menjadi “< 2 Jam” = 0, “2 – 5 Jam” = 1 dan “>5 Jam” = 2.

Penerapan Algoritma C4.5

Penerapan algoritma C4.5 pada penelitian ini dilakukan untuk membentuk model klasifikasi yang mampu memprediksi kategori prestasi akademik siswa berdasarkan atribut kebiasaan harian. Proses penghitungan dimulai dengan menentukan nilai awal dari keseluruhan dataset. Pada dataset penelitian ini terdapat dua kelas yaitu ≥ 85 dan 70–84. Jumlah data pada kelas ≥ 85 adalah 60 siswa dan jumlah data pada kelas 70–84 adalah 20 siswa, sehingga total data berjumlah 80 siswa. Nilai entropi total dihitung menggunakan rumus:

$$\text{Entropy}(S) = - \left(\frac{60}{80} \log 2 \frac{60}{80} \right) - \left(\frac{20}{80} \log 2 \frac{20}{80} \right)$$

$$\text{Entropy}(S) = - (0.75 \log 2 0.75) - (0.25 \log 2 0.25)$$

$$\text{Entropy}(S) = -(0.75x - 0.554585) - (0.25x - 2)$$

$$\text{Entropy}(S) = 0.41594 + 0.5 = 0.91594$$

Nilai entropi 0,91594 ini digunakan sebagai dasar untuk menghitung gain dari setiap atribut. Atribut pertama yang dianalisis adalah Durasi Belajar yang telah melalui proses encoding menjadi tiga kategori yaitu 0 = < 1 Jam, 1 = 1–2 Jam dan 2 = > 2 Jam. Jumlah siswa pada kategori < 1 Jam adalah 35 orang dengan distribusi 17 siswa pada kelas “ ≥ 85 ” dan 18 siswa pada kelas “70–84”. Nilai entropinya dihitung sebagai berikut:

$$\begin{aligned}\text{Entropy}(< 1 \text{ Jam}) &= - \left(\frac{17}{35} \log 2 \frac{17}{35} \right) - \left(\frac{18}{35} \log 2 \frac{18}{35} \right) \\ &= -(0.48571 \log 2 0.48571) - (0.51429 \log 2 0.51429) \\ &= -(0.48571x - 1.04015) - (0.51429x - 0.95800) \\ &= 0.50556 + 0.49268 = 0.998\end{aligned}$$

Perhitungan yang sama diterapkan pada kategori 1–2 Jam dan > 2 Jam. Misalkan nilai entropi kategori 1–2 Jam adalah 0,881 dan kategori > 2 Jam adalah 0,722. Maka gain untuk atribut Durasi Belajar diperoleh dari:

$$\begin{aligned}\text{Gain}(\text{Durasi Belajar}) &= 0.998 - \left(\frac{35}{80}x 0.998 + \frac{25}{80}x 0.881 + \frac{10}{80}x 0.722 \right) \\ \text{Gain} &= 0.998 - (0.436 + 0.275 + 0.090) \\ &= 0.998 - 0.801 = 0.197\end{aligned}$$

Proses perhitungan dilanjutkan pada atribut Durasi Tidur dan Durasi Penggunaan Gadget. Untuk setiap atribut dilakukan perhitungan entropi per kategori, kemudian dihitung nilai gain dan terakhir dihitung gain ratio untuk menentukan atribut terbaik sebagai akar pohon

keputusan. Atribut Durasi Tidur menghasilkan gain sebesar 0,215 dan atribut Penggunaan Gadget menghasilkan gain sebesar 0,167. Nilai-nilai tersebut kemudian dibagi dengan split info masing-masing atribut untuk memperoleh gain ratio. Gain ratio Durasi Belajar = 0,197, Durasi Tidur = 0,102, dan Penggunaan Gadget = 0,054. Dengan demikian, atribut Durasi Tidur menjadi atribut terbaik untuk dijadikan akar pohon keputusan karena memiliki gain ratio tertinggi.

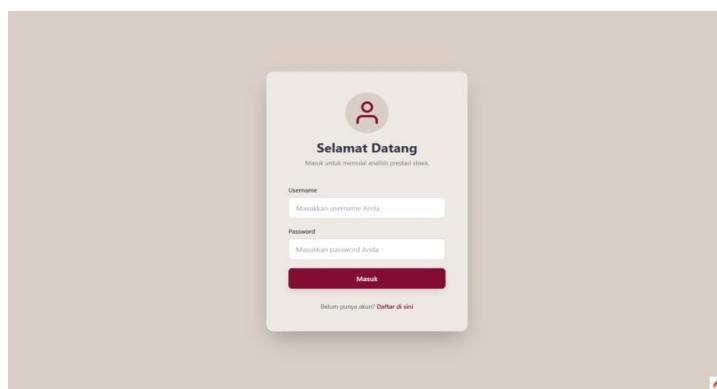
Implementasi Sistem

Tahap implementasi sistem merupakan proses realisasi dari seluruh rancangan yang telah disusun pada tahap analisis dan perancangan. Implementasi dilakukan untuk mengubah desain konseptual menjadi sebuah sistem yang dapat dijalankan dan digunakan oleh pengguna. Pada tahap ini setiap komponen sistem diwujudkan dalam bentuk fungsional sesuai dengan alur proses yang telah ditetapkan, sehingga sistem dapat beroperasi secara optimal dalam melakukan klasifikasi prestasi akademik siswa.

Dalam penelitian ini sistem dikembangkan dalam bentuk aplikasi berbasis web yang dirancang untuk mengelola data siswa serta menjalankan proses klasifikasi secara otomatis menggunakan algoritma C4.5. Aplikasi ini menyediakan beberapa fitur utama, seperti halaman input data siswa, pengelolaan dataset, proses analisis klasifikasi, penyajian hasil klasifikasi serta halaman riwayat yang menampilkan catatan hasil analisis yang telah dilakukan.

Tampilan Halaman Login

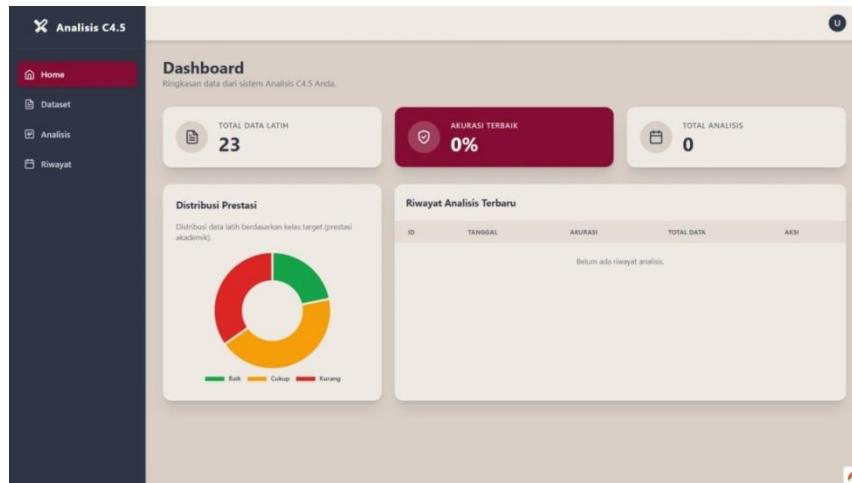
Tampilan halaman login merupakan antarmuka awal yang diakses oleh pengguna sebelum memasuki sistem.



Gambar 2. Tampilan Halaman Login

Tampilan Halaman Utama

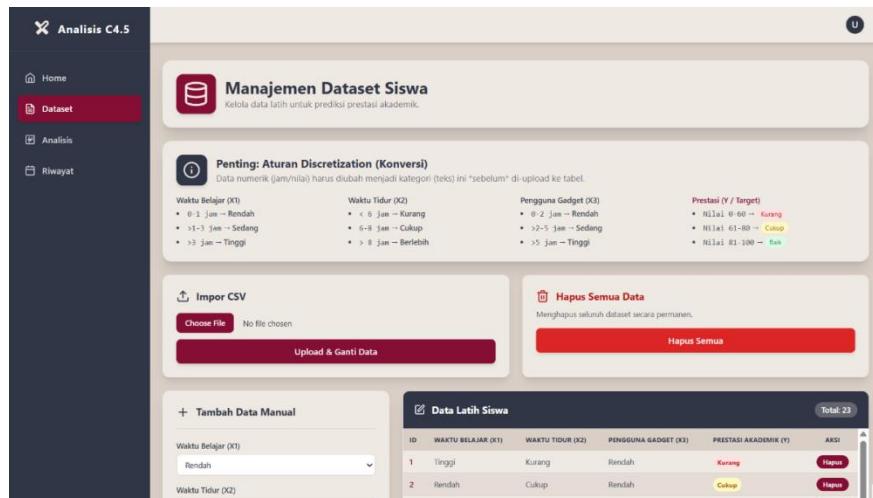
Tampilan halaman utama merupakan antarmuka yang ditampilkan setelah pengguna berhasil melakukan proses login. Halaman ini berfungsi sebagai pusat navigasi yang memberikan akses cepat menuju seluruh fitur yang tersedia pada sistem klasifikasi prestasi akademik.



Gambar 3. Tampilan Halaman Utama

Tampilan Halaman Dataset

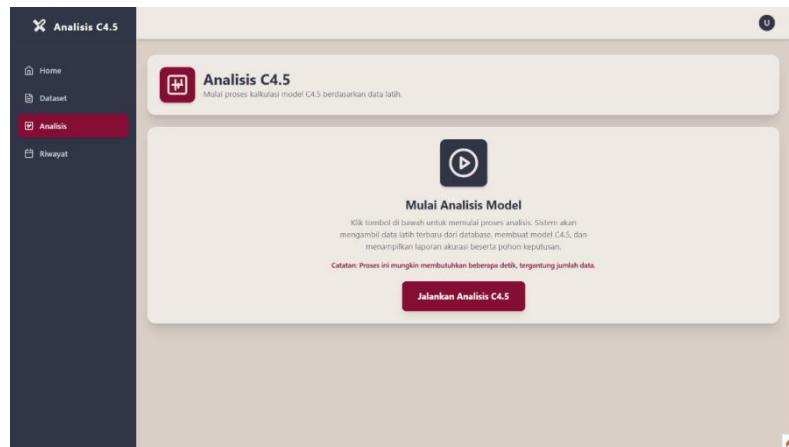
Tampilan halaman dataset merupakan antarmuka yang digunakan untuk menampilkan dan mengelola seluruh data siswa yang menjadi dasar dalam proses klasifikasi prestasi akademik.



Gambar 4. Tampilan Halaman Dataset

Tampilan Halaman Analisis

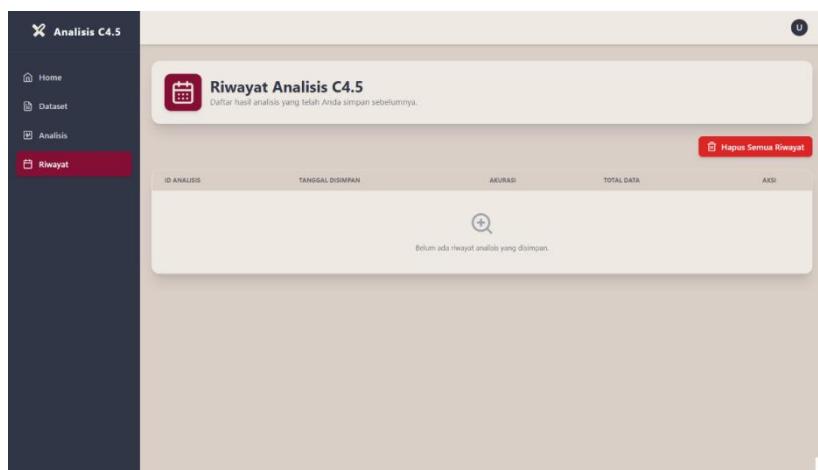
Tampilan halaman analisis merupakan antarmuka yang digunakan oleh pengguna untuk menjalankan proses klasifikasi prestasi akademik siswa menggunakan algoritma C4.5.



Gambar 5. Tampilan Halaman Analisis

Tampilan Halaman Riwayat

Tampilan halaman riwayat merupakan antarmuka yang menampilkan daftar hasil klasifikasi prestasi akademik yang telah dilakukan sebelumnya. Halaman ini berfungsi untuk menyimpan dan menampilkan seluruh catatan analisis yang telah diproses oleh sistem sehingga pengguna dapat melihat kembali kategori prestasi yang dihasilkan oleh algoritma C4.5.

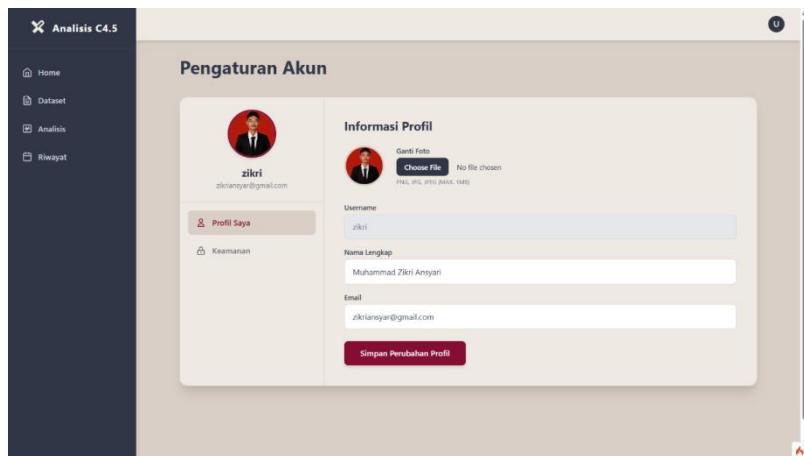


Gambar 6. Tampilan Halaman Riwayat

Tampilan Halaman Akun

Tampilan halaman akun merupakan antarmuka yang digunakan untuk menampilkan informasi pengguna yang terdaftar dalam sistem serta menyediakan fitur untuk mengelola data akun. Pada halaman ini pengguna dapat melihat data pribadi seperti nama pengguna, username

dan email. Informasi tersebut ditampilkan dalam format yang sederhana dan terstruktur agar mudah dipahami.



Gambar 7. Tampilan Halaman Akun

5. KESIMPULAN

Penelitian ini berhasil merancang dan membangun sistem klasifikasi prestasi akademik berbasis web melalui analisis kebutuhan, perancangan dengan diagram UML, dan pembuatan rancangan antarmuka, sehingga menghasilkan sistem yang terstruktur, mudah dipahami, dan mampu mendukung pengelolaan data siswa. Algoritma C4.5 berhasil diterapkan sebagai metode utama untuk membentuk aturan klasifikasi berdasarkan variabel durasi belajar, durasi tidur, dan durasi penggunaan gadget. Sistem yang dikembangkan mampu menghasilkan kategori prestasi akademik secara otomatis berdasarkan input pengguna, dan hasil pengujian menunjukkan bahwa sistem bekerja secara optimal dalam mengolah data, menjalankan proses klasifikasi, serta menampilkan hasil yang akurat dan mudah dipahami, sehingga dapat menjadi alat bantu evaluasi untuk menilai kecenderungan prestasi siswa berdasarkan kebiasaan belajar dan gaya hidup sehari-hari.

SARAN

Berdasarkan proses penelitian, perancangan, dan implementasi sistem klasifikasi prestasi akademik yang telah dilakukan, beberapa saran pengembangan yang dapat diterapkan di masa mendatang antara lain penambahan variabel pendukung seperti tingkat kehadiran, metode belajar, dan nilai akademik lainnya untuk meningkatkan akurasi model, peningkatan antarmuka dan pengalaman pengguna melalui penyajian hasil dalam bentuk grafik atau statistik agar pola klasifikasi lebih mudah dipahami, serta pengembangan sistem menjadi aplikasi web yang dapat

diakses secara online maupun terintegrasi dengan perangkat mobile sehingga analisis data dan hasil klasifikasi dapat diperoleh secara lebih fleksibel dan real-time.

DAFTAR REFERENSI

- Anisa, A., Magfirah, N., & Thahir, R. (2020). Peranan self efficacy dan self regulated learning terhadap prestasi akademik mahasiswa. *Biodik*, 7(2), 63–70. <https://doi.org/10.22437/bio.v7i2.12824>
- Arief, S. F., & Sugiarti, Y. (2022). Literature review: Analisis metode perancangan sistem informasi akademik berbasis web. *Jurnal Ilmiah Ilmu Komputer*, 8(2), 87–93. <https://doi.org/10.35329/jiik.v8i2.229>
- Bachtiar, L., & Mahradianur, M. (2023). Analisis data mining menggunakan metode algoritma C4.5 menentukan penerima bantuan langsung tunai. *Jurnal Informatika*, 10(1), 28–36. <https://doi.org/10.31294/inf.v10i1.15115>
- Firdaus, N. F. 'Aini, Midyanti, D. M., & Hidayati, R. (2023). Penerapan metode C4.5 untuk klasifikasi potensi akademik siswa (Studi kasus: SMP DDI Sultan Syarif Abdurachman Pontianak). *Coding: Jurnal Komputer dan Aplikasi*, 11(1), 92–101. <https://doi.org/10.26418/coding.v11i1.54266>
- Gusnawan, R., Wijayanto, P., & Rosely, E. (2019). Aplikasi kelompok informasi masyarakat (KIM) berbasis web. *E-Proceeding of Applied Science*, 5(2), 1123–1132.
- Hendita, G., & Kusuma, A. (2021). Perancangan skema sistem keamanan jaringan web server menggunakan web application firewall dan Fortigate untuk mencegah kebocoran data di masa pandemi Covid-19. *Zeitschrift Für Wirtschaftlichen Fabrikbetrieb*, 92(1–2), 37. <https://doi.org/10.1515/zwf-1997-921-220>
- Hidayat, A., Yani, A., Rusidi, & Saadulloh. (2019). Membangun website SMA PGRI Gunung Raya Ranau menggunakan PHP dan MySQL. *JTIM: Jurnal Teknik Informatika Mahakarya*, 2(2), 41–52.
- Husein, A. M., & Hutaikur, R. E. H. (2022). Penerapan algoritma C4.5 dalam pemilihan siswa berprestasi di SMPN 10 Medan. *Digital Transformation Technology*, 2(1), 8–11. <https://doi.org/10.47709/digitech.v2i1.1768>
- Indriati, K., Muchlas, M., & Sayuti, M. (2023). Kebiasaan belajar siswa sekolah menengah kejuruan saat pandemi Covid-19 di SMK Muhammadiyah Purwodadi Purworejo. *JIM: Jurnal Ilmiah Mahasiswa*, 8(3). <https://jim.usk.ac.id/sejarah/article/view/25117>
- Kasus, S., Karunia, U. D., & Utama, S. (2021). Analisis postur kerja pada stasiun kerja proses som kaos kaki. *Journal of Industrial Engineering and Technology (Jointech)*, 1(2), 52–60. [https://doi.org/10.24176/jointech.v1i2.6500](https://doi.org/10.24176/jointtech.v1i2.6500)
- Maydianto, & Ridho, M. R. (2021). Rancang bangun sistem informasi point of sale dengan framework CodeIgniter pada CV Powershop. *Jurnal Comasie*, 2(1), 50–59.
- Nurhalizah, R. S., Ardianto, R., & Purwono, P. (2024). Analisis supervised dan unsupervised learning pada machine learning: Systematic literature review. *Jurnal Ilmu Komputer dan Informatika*, 4(1), 61–72. <https://doi.org/10.54082/jiki.168>
- Partogi, Y., & Pasaribu, A. (2022). Perancangan metode decision tree terhadap sistem perpustakaan STMIK Kuwera. *Jurnal Sistem Informasi dan Teknologi (SINTEK)*, 1(2), 20–25. <https://doi.org/10.56995/sintek.v1i2.4>

- Putra, D. W. T., & Andriani, R. (2019). Unified Modelling Language (UML) dalam perancangan sistem informasi permohonan pembayaran restitusi SPPD. *Jurnal TeknoIf*, 7(1), 32–39. <https://doi.org/10.21063/jtif.2019.v7.1.32-39>
- Rahmat Inggi, & Alam, H. P. (2023). Analisis forensik web browser pada perangkat Android. *Simtek: Jurnal Sistem Informasi dan Teknik Komputer*, 8(1), 215–220. <https://doi.org/10.51876/simtek.v8i1.249>
- Rahmi, E. R., Yumami, E., & Hidayasari, N. (2023). Analisis metode pengembangan sistem informasi berbasis website: Systematic literature review. *Remik*, 7(1), 821–834. <https://doi.org/10.33395/remik.v7i1.12177>
- Sukri, M. H., & Handrianto, Y. (2024). Penerapan algoritma C4.5 dalam menentukan prediksi prestasi siswa pada SMPN 51 Jakarta. *Informatics and Computer Engineering Journal*, 4(1), 11–24. <https://doi.org/10.31294/icej.v4i1.2582>
- Tukino, T. (2019). Penerapan algoritma C4.5 untuk memprediksi keuntungan pada PT SMOE Indonesia. *Jurnal Sistem Informasi Bisnis*, 9(1), 39–46. <https://doi.org/10.21456/vol9iss1pp39-46>