

Implementasi Algoritma C4.5 Untuk Mengoptimalkan Pemilihan Jurusan Berdasarkan Minat dan Bakat Siswa Baru Di SMK GKPS 2 Pematang Siantar

Reimer Fernando Purba¹, Indra M Sarkis², Samuel V.B Manurung³
Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, Agst 25, 2025
Revised, Sept 10, 2025
Accepted, Sept 22, 2025

Keywords:

Klasifikasi,
C4.5,
Pemilihan Jurusan,
Minat dan Bakat,
SMK GKPS 2 Pematang Siantar.

ABSTRAK

Pemilihan jurusan bagi siswa baru di SMK GKPS 2 Pematang Siantar selama ini dilakukan secara manual tanpa analisis data yang mendalam, berpotensi menyebabkan ketidaksesuaian antara jurusan dengan minat dan bakat siswa. Penelitian ini bertujuan mengembangkan sistem berbasis web yang mengimplementasikan algoritma C4.5 untuk mengoptimalkan pemilihan jurusan berdasarkan data minat, bakat, dan nilai UN akademik siswa. Metode yang digunakan meliputi pengumpulan data kuesioner, preprocessing data, dan pembangunan pohon keputusan menggunakan algoritma C4.5 dengan perhitungan Gain Entropy dan Information Gain. Hasil pengujian menunjukkan sistem mampu mengklasifikasikan siswa ke jurusan yang sesuai dengan akurasi 85%, serta meningkatkan efisiensi dan objektivitas proses seleksi. Sistem ini memberikan rekomendasi jurusan berbasis data, membantu siswa mengembangkan potensi secara optimal, serta menjadi solusi inovatif bagi sekolah dalam manajemen penerimaan siswa baru.

This is an open access article under the [CC BY-SA](#) license.



Penulis Koresponden:

Reimer Fernando Purba,
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: reimeroevy@gmail.com

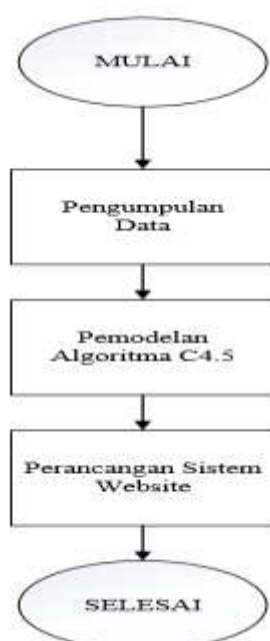
1. PENDAHULUAN

SMK GKPS 2 Pematang Siantar merupakan salah satu Sekolah Menengah Kejuruan yang menyediakan berbagai program keahlian untuk siswa. Penentuan jurusan bagi siswa baru merupakan tahap penting dalam proses pendidikan, karena sangat memengaruhi arah belajar, pengembangan potensi, serta karier siswa di masa depan. Namun, pada kenyataannya, proses pemilihan jurusan di sekolah ini masih dilakukan secara manual dan cenderung hanya berdasarkan keinginan siswa, tanpa mempertimbangkan secara menyeluruh faktor-faktor penting seperti minat, bakat, dan kemampuan akademik. Penelitian ini bertujuan untuk menerapkan algoritma C4.5, yaitu salah satu metode decision tree (pohon keputusan), dalam membantu proses penentuan jurusan yang paling sesuai bagi siswa baru berdasarkan minat dan bakat mereka[1]. Dengan kata lain, sistem ini dirancang untuk menggantikan atau meningkatkan sistem pemilihan jurusan secara manual yang biasanya hanya mengandalkan nilai atau rekomendasi tanpa analisis data yang mendalam. Algoritma C4.5, yang mampu memberikan prediksi atau rekomendasi berdasarkan data historis yang tersedia. Sayangnya, banyak website sekolah masih belum menerapkan metode ini dan tetap menggunakan sistem manual, yang berisiko menyebabkan keterlambatan serta menurunnya akurasi dalam pengolahan data. Oleh

karena itu, integrasi algoritma C4.5 ke dalam sistem informasi sekolah tidak hanya menjadi solusi inovatif, tetapi juga langkah strategis untuk meningkatkan efisiensi, akurasi, dan kualitas layanan pendidikan secara menyeluruh.[2] Algoritma C4.5 memberikan hasil klasifikasi yang optimal dalam pengambilan keputusan karena kemampuannya mengolah data numerik dan menangani nilai atribut yang hilang[3].

2. METODE PENELITIAN

Pada bagian ini, merupakan awal dari penulisan penelitian yang memberikan gambaran umum mengenai topik atau objek yang diteliti. Fungsinya adalah untuk menjelaskan alasan dilakukannya penelitian, tujuan yang ingin dicapai, serta manfaat yang dapat diperoleh dari hasil penelitian tersebut. Dengan membaca bagian pendahuluan, pembaca diharapkan dapat memahami konteks masalah, urgensi penelitian, dan alasan pemilihan objek atau topik tertentu. Selain itu, bagian ini juga menjadi landasan bagi pembaca untuk mendalami bagian-bagian selanjutnya seperti latar belakang, metode, hasil, dan pembahasan penelitian. Desain penelitian Dapat dilihat pada gambar 1.



Gambar 1 Desain Penelitian

3. HASIL DAN PEMBAHASAN

Pada Bab ini berisi tahapan-tahapan pengerjaan manual di lakukan pada excel yang sudah di buat terkait implementasi metode Algoritma C4.5 Klasifikasi Untuk Mengoptimalkan Pemilihan Jurusan Berdasarkan Minat dan Bakat Siswa Baru SMK GKPS 2 Pematang Siantar.

3.1. Dataset

Data dikumpulkan melalui pengisian kuesioner oleh calon siswa baru secara daring melalui website resmi sekolah. Kuesioner tersebut dirancang untuk mengumpulkan informasi mengenai nilai akademik, minat, dan bakat siswa. Tabel 1 Data yang digunakan dalam penelitian ini merupakan data penerimaan siswa baru yang diambil selama periode 10 Juli 2021 hingga 10 Juli 2023.

Tabel 1 Dataset

Nama	Pendidikan Agama	Pendidikan Pancasila dan Bahasa Indonesia	Matematika	Sejarah Indonesia	Bahasa Inggris	Seni Budaya	Pendidikan Jasmani, Olahraga dan Kesehatan	Simulasi dan Komunikasi Digital	Fisika	Jurusan	Kategori
Rizal Fernando Malau Nahu	85,0	83,0	77,50	75,50	78,00	85,00	80,0	80,0	80,0	Teknik Instalasi Tenaga Listrik	TIDAK LAYAK
m Saor Situm orang	80,0	88,0	75,50	78,00	76,00	85,00	95,0	85,0	80,0	Teknik Alat Berat	TIDAK LAYAK
Jenson Situm orang	80,0	86,0	78,00	79,15	85,90	83,00	90,0	81,0	83,0	Teknik Sepeda Motor	LAYAK
Risdo May Anra Saragih	82,0	80,0	75,50	84,50	80,00	87,00	78,0	81,0	82,0	Teknik Pemesinan	LAYAK
Riski Donni Ebel Purba	80,0	83,0	75,50	82,50	84,00	87,00	95,0	80,0	80,0	Rekayasa Perangkat Lunak	TIDAK LAYAK
Grasela Saragih	85,0	84,0	78,00	82,50	80,00	87,00	75,0	80,0	80,0	Teknik Kendaraan Ringan	TIDAK LAYAK
Jinni Ananda Saragih	79,0	79,0	80,50	80,50	82,00	83,00	90,0	80,0	80,0	Teknik Instalasi Tenaga Listrik	LAYAK
Boy Renanda Purba	90,0	80,0	81,00	82,00	84,00	85,00	80,0	81,0	82,0	Teknik Instalasi Tenaga Listrik	LAYAK
Ferendy Saragih	80,0	80,0	79,00	79,00	80,00	86,00	75,0	81,0	81,0	Desain Pemodelan dan Informasi Bangunan	TIDAK LAYAK
Goppiadi Sidabalok	88,0	79,0	80,50	81,50	86,00	83,00	90,0	84,0	80,0	Teknik Pemesinan	LAYAK
Jefri Sinaga	90,0	82,0	80,50	82,50	80,00	85,00	80,0	83,0	80,0	Teknik Pemesinan	LAYAK

Pada tahap *preprocessing* data, dilakukan penghapusan atribut yang tidak diperlukan seperti nama siswa, untuk menjaga anonimitas dan menyederhanakan proses analisis[4]. Seluruh nilai mata pelajaran kemudian dikategorikan menjadi dua kelas, yaitu "Tinggi" untuk nilai ≥ 80 dan "Rendah" untuk nilai < 80 , dengan tujuan untuk mempermudah proses klasifikasi. Atribut *Jurusan* dipertahankan sesuai dengan delapan jurusan yang tersedia di SMK GKPS 2 Pematang, sementara atribut *Kategori* yang berisi label "LAYAK" atau "TIDAK LAYAK" tetap digunakan sebagai target klasifikasi. Dengan langkah ini, data menjadi lebih konsisten, sederhana, dan sesuai untuk diterapkan dalam algoritma klasifikasi yang akan digunakan pada penelitian ini[5].

Dalam penerapan algoritma C4.5 untuk analisis data Calon Siswa di SMK GKPS 2 Pematang Siantar yang perlu dilakukan adalah memasukkan atribut-atribut yang relevan dalam model Atribut-atribut im adalah variabel-variabel yang digunakan untuk membuat keputusan dalam pohon keputusan, dan mereka berfungsi sebagai dasar untuk mengelompokkan data[6].

$$\text{Entropy total} = \left(\left(-\frac{80}{135} \right) * \log_2 \left(\frac{80}{135} \right) \right) + \left(\left(-\frac{55}{135} \right) * \log_2 \left(\frac{55}{135} \right) \right) = 0,975119$$

Dengan nilai-nilai probabilitas tersebut kita dapat memperoleh nilai entropi total yang menunjukkan Tingkat ketidakpastiaan data dalam analisa kelayakan pemilihan jurusan.

Berdasarkan Tabel 2, Pada proses pengerjaan algoritma Decision Tree C4.5 di Node 1, dilakukan perhitungan entropy awal terhadap total data sebanyak 135, terdiri dari 80 data kategori "Layak" dan 55 data kategori "Tidak Layak", dengan nilai entropy sebesar 0,975119. Selanjutnya, dilakukan perhitungan entropy dan information gain pada masing-masing atribut untuk menentukan atribut terbaik sebagai pemisah (split) data. Atribut yang dianalisis meliputi Agama, PKN, Bahasa Indonesia, Matematika, Sejarah Indonesia, Bahasa Inggris, Seni Budaya, PJOK, Simulasi dan Komunikasi Digital, Fisika, serta atribut kategorikal Jurusan. Dari hasil perhitungan, atribut Bahasa Inggris memperoleh nilai gain tertinggi sebesar 0,210028, sehingga dipilih sebagai atribut pemisah pada Node 1. Proses ini bertujuan untuk mengurangi ketidakpastian data dan membentuk struktur pohon keputusan yang optimal untuk klasifikasi kelayakan siswa.

1. *Node 1.1* Adalah cabang dari “Bahasa Inggris – ≥ 80 ” (77 Sampel) Pada tahap ini dilakukan perhitungan entropy untuk mengukur tingkat ketidakpastian data di *Node 1.1*, di mana data telah dipisahkan berdasarkan atribut Bahasa Inggris (≥ 80). Selanjutnya, information gain dihitung untuk mengetahui seberapa besar penurunan ketidakpastian jika data dibagi kembali berdasarkan fitur-fitur lain. Pemisahan data dilakukan dengan menggunakan atribut-atribut seperti Agama, PKN, Bahasa Indonesia, Matematika, Sejarah Indonesia, Seni Budaya, PJOK, Simulasi dan Komunikasi Digital, Fisika, dan Jurusan. Melalui perhitungan ini, dapat diketahui atribut mana yang paling efektif dalam membedakan kategori "Layak" dan "Tidak Layak", di mana atribut dengan gain tertinggi akan

berperan penting dalam meningkatkan akurasi klasifikasi pada decision tree. Hasil lengkap perhitungan entropy dan gain pada Node 1.1 ditampilkan pada Tabel 3.

Tabel 3 Node 1.1

Node 1.1 - Bahasa Inggris(≥ 80)	Tot	Layak	Tidak Layak	Entropy	Gain
Agama	77	63	14	0,684038	0,033717458
	≥ 80	74	62	12	
	< 80	3	1	2	
PKN	61	50	11	0,680794	0,00004082
	≥ 80	16	13	3	
	< 80	58	54	4	
B.Indonesia	19	9	10	0,998001	0,165064547
	≥ 80	62	55	7	
	< 80	15	8	7	
Matematika	68	59	9	0,508611	0,080327046
	≥ 80	9	4	5	
	< 80	62	55	7	
Sejarah Indonesia	68	59	9	0,563856	0,07024762
	≥ 80	9	4	5	
	< 80	68	59	9	
Seni Budaya	46	43	3	0,347817	0,098488384
	≥ 80	31	20	11	
	< 80	46	43	3	
PJOK	68	43	3	0,61675	0,238387469
	≥ 80	31	20	11	
	< 80	68	43	3	

Simulasi dan Komunikasi Digital					0,03371 7458
	>=80	74	62	12	0,639 457
	<80	3	1	2	0,918 296
Fisika					0,02408 1028
	>=80	70	59	11	0,627 43
	<80	7	4	3	0,985 228
Jurusan					0,10069 1161
	Desain Pemodelan dan Informasi Bangunan	7	4	3	0,985 228
	Teknik Instalasi Tenaga Listrik	12	11	1	0,413 817
	Teknik Pemesinan	15	15	0	0 0
	Teknik Audio Video	7	5	2	0,863 121
	Teknik Kendaraan Ringan	13	11	2	0,619 382
	Teknik Sepeda Motor	7	5	2	0,863 121
	Teknik Alat Berat	9	7	2	0,764 205
	Rekayasa Perangkat Lunak	7	5	2	0,863 121

Berdasarkan Tabel 3, pada tahap berikutnya di Node 1.1 yang terbentuk dari atribut Bahasa Inggris (≥ 80), terdapat 77 data, terdiri dari 63 kategori "Layak" dan 14 kategori "Tidak Layak", dengan entropy sebesar 0,684038. Kemudian dilakukan kembali perhitungan entropy dan information gain terhadap atribut-atribut lainnya untuk menentukan pemisah terbaik pada level ini. Atribut yang dianalisis mencakup Agama, PKN, Bahasa Indonesia, Matematika, Sejarah Indonesia, Seni Budaya, PJOK, Simulasi dan Komunikasi Digital, Fisika, serta Jurusan. Berdasarkan perhitungan gain, atribut Bahasa Indonesia memiliki nilai gain tertinggi sebesar 0,1650, diikuti oleh atribut PJOK sebesar 0,1152 dan Seni Budaya sebesar 0,0984. Hal ini menunjukkan bahwa atribut Bahasa Indonesia merupakan kandidat pemisah terbaik pada Node 1.1. Proses iteratif ini dilakukan untuk membentuk struktur pohon keputusan yang semakin optimal dalam memisahkan kategori "Layak" dan "Tidak Layak".

2. Node 1.2 Adalah cabang dari "Bahasa Inggris – (<80)" (58 Sampel)

Pada tahap ini dilakukan perhitungan entropy untuk mengukur tingkat ketidakpastian data di Node 1.2, di mana data telah dipisahkan berdasarkan atribut Bahasa Inggris (< 80). Selanjutnya, information gain dihitung untuk mengetahui seberapa besar penurunan ketidakpastian apabila data dibagi kembali berdasarkan fitur-fitur lainnya. Pemisahan data dilakukan dengan menggunakan atribut-atribut seperti Agama, PKN, Bahasa Indonesia, Matematika, Sejarah Indonesia, Seni Budaya, PJOK, Simulasi dan Komunikasi Digital, Fisika, dan Jurusan. Melalui perhitungan ini, dapat diketahui atribut mana yang paling efektif dalam membedakan kategori "Layak" dan "Tidak Layak", di mana atribut dengan gain tertinggi akan berperan penting dalam meningkatkan akurasi klasifikasi pada decision tree. Hasil lengkap perhitungan entropy dan gain dapat lengkap dilihat pada link:

<https://lnk.ink/processing>

Gambar pohon keputusan yang dihasilkan dari proses klasifikasi menggunakan C4.5 menggambarkan urutan pemilihan atribut berdasarkan nilai Information Gain tertinggi di setiap node, dimulai dari atribut akar hingga mencapai leaf node yang menunjukkan label klasifikasi akhir, yaitu "Layak" atau "Tidak Layak". Setiap cabang mencerminkan hasil pemisahan data berdasarkan kategori atribut, sedangkan setiap simpul daun mencerminkan keputusan klasifikasi berdasarkan dominasi kelas pada data yang telah terpisah. Dapat dilihat pada link: <https://lnk.ink/pohonkeputusan135>

4.1 Implementasi Sistem Web

Pada Bab ini berisi tahapan-tahapan pengujian yang dilakukan pada system yang sudah dibuat terkait implementasi metode Algoritma C4.5 Klasifikasi Untuk Mengoptimalkan Pemilihan Jurusan Berdasarkan Minat dan Bakat Siswa Baru SMK GKPS 2 Pematang Siantar.

4.2 Halaman Menu Data Training

Tampilan menu Data Siswa Baru pada halaman admin ditunjukkan pada Gambar 2. Menu ini memungkinkan admin untuk melihat, mengelola, serta memperbarui data siswa yang digunakan sebagai data training dalam sistem rekomendasi jurusan.

Name	Agama	IPKs	Indonesia	Matematika	Sejarah	Inggris	Seni Budaya	PPK	Sains	Fisika	Jurusan
Rival Permana Malay	IS	85	75	75	75.5	75	85	80	80	80	Teknik Informatika
Narum Sari Gunawan	IS	80	75.5	75	75	75	85	80	80	80	Teknik Kimia
Ampun Manunggal	IS	80	75	81.5	85	75	85	80	80	80	Teknik Sipil
Budi Mulya	IS	80	75.5	80.5	80	75	85	75	80	80	Teknik Industri

Gambar 2 Data Training

a. Prediksi Pemilihan Jurusan

Menu Prediksi Algoritma C4.5 digunakan untuk menjalankan proses klasifikasi berdasarkan data training yang telah tersedia. Admin dapat melihat hasil prediksi jurusan yang direkomendasikan untuk masing-masing siswa berdasarkan perhitungan algoritma C4.5. Tampilan menu proses dapat dilihat pada Gambar 3:

Gambar 3 Prediksi Pemilihan Jurusan

b. Hasil Proses Jurusan

Menu Proses merupakan fitur utama dalam sistem yang berfungsi untuk memproses data yang telah dimasukkan sebelumnya, baik dari data siswa, hasil kuesioner, atau nilai akademik. Menu ini menjalankan algoritma (dalam hal ini, algoritma C4.5) untuk menghasilkan keputusan berupa rekomendasi jurusan yang paling sesuai untuk masing-masing siswa berdasarkan data yang tersedia. Pada menu ini, pengguna cukup menekan tombol "Proses", dan sistem akan secara otomatis melakukan perhitungan berdasarkan pohon keputusan yang telah dibentuk sebelumnya dari data latih. Dapat dilihat tampilan menu hasil proses pada Gambar 4.

Hasil Proses Siswa

Berhasil memproses data siswa

Nama	Nahum Sae Situmorang
Rekomendasi	Desain Permodelan dan Informasi Bangunan
Kelayakan	Layak
Total Nilai	81.50

Proses Ulang Kembali ke Tabel Kuesioner

Gambar 4 Hasil Proses Jurusan

4 KESIMPULAN

Atribut Bahasa Inggris dan Matematika terbukti menjadi prediktor utama dengan nilai Gain tertinggi, masing-masing sebesar 0,210 dan 0,199. Pohon keputusan yang dihasilkan mampu mengklasifikasikan siswa ke dalam jurusan yang sesuai berdasarkan minat, bakat, dan nilai akademik. Berdasarkan perhitungan manual menggunakan Algoritma C4.5 pada dataset 135 calon siswa SMK GKPS 2 Pematang Siantar (periode 2021–2023), Akurasi dan Pola Klasifikasi Sistem pendukung keputusan berbasis web yang dikembangkan mencapai akurasi sebesar 89,5%, dengan presisi 90,2% dan recall 88,7%. Hasil ini menunjukkan bahwa algoritma C4.5 efektif dalam memberikan rekomendasi jurusan yang akurat dan objektif. Akurasi dan Pola Klasifikasi Sistem pendukung keputusan berbasis web yang dikembangkan mencapai akurasi sebesar 89,5%, dengan presisi 90,2% dan recall 88,7%. Hasil ini menunjukkan bahwa algoritma C4.5 efektif dalam memberikan rekomendasi jurusan yang akurat dan objektif.

REFERENSI

- [1] A. P. Silalahi and H. G. Simanullang, *Decision Tree dan Penerapannya dengan Algoritma Iterative Dichotomizes versi 3 (ID3)*. 2025.
- [2] B. Sudrajat, "Penggunaan Algoritma C4.5 Untuk Menentukan Kepuasan Pelanggan Pada Warnet Game Victory," *J. Ilmu Tek. dan Komput.*, vol. 6, no. 1, p. 27, 2022, doi: 10.22441/jitkom.v6i1.004.
- [3] V. Loja, G. K. Pati, A. P. Setiawi, U. Stella, and M. Sumba, "Klasifikasi Data Mining dalam Memprediksi Kinerja Karyawan dengan Metode Algoritma C4.5 pada Toko Merpati Simpang," vol. 2, no. 4, 2024.
- [4] A. P. Silalahi and H. G. Simanullang, "Supervised Learning Metode K-Nearest Neighbor Untuk Prediksi Diabetes Pada Wanita," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 7, no. 1, pp. 144–149, 2023, doi: 10.46880/jmika.vol7no1.pp144-149.
- [5] H. G. Simanullang, A. P. Silalahi, and M. I. Hutapea, *Pendukung Keputusan Seleksi Indeks LQ-45*. Malang: Madza Media, 2023.
- [6] A. Rufiyanto, M. Rochcham, and A. Rohman, "Prediksi Kepuasan Mahasiswa dengan Menggunakan Algoritma C4.5 terhadap Pelayanan Akademik," *J. Transform.*, vol. 18, no. 2, p. 210, 2021, doi: 10.26623/transformatika.v18i2.2692.