



# Sistem Rekomendasi Jurusan Pendidikan Menggunakan Algoritma C4.5 Berbasis Web untuk Siswa Sekolah Menengah Atas

Edy<sup>1</sup>, Rino<sup>2</sup>, Desiyanna Lasut<sup>3</sup>

<sup>1,2</sup> Universitas Buddhi Dharma, Teknik Informatika, Tangerang, Indonesia

## SUBMISSION TRACK

Received: 31 Agustus 2024  
Final Revision: March 26, 2025  
Available Online: March 28, 2025

## KEYWORD

Kepribadian, Algoritma C4.5, Tree Decision, Jurusan, Rekomendasi, Data Mining

## KORESPONDENSI

Phone: 082119466523  
E-mail: [desiyanna\\_lasut@yahoo.com](mailto:desiyanna_lasut@yahoo.com)

## ABSTRACT

Pemilihan jurusan pendidikan yang sesuai dengan potensi diri sangat penting bagi siswa sekolah menengah atas sebagai dasar penentuan arah karier dan studi lanjut. Penelitian ini mengembangkan sistem berbasis web untuk merekomendasikan jurusan pendidikan menggunakan algoritma C4.5. Algoritma ini dipilih karena mampu mengklasifikasikan data secara efisien dan menghasilkan pohon keputusan yang mudah dipahami. Data yang digunakan meliputi hasil tes minat, nilai akademik, dan preferensi siswa. Sistem ini dirancang agar siswa dapat mengetahui jurusan yang paling sesuai dengan profil mereka secara cepat dan akurat. Hasil pengujian menunjukkan bahwa sistem memiliki tingkat akurasi antara 90–95%, serta mampu memberikan rekomendasi secara real-time berbasis data input. Sistem ini dapat membantu pihak sekolah dan siswa dalam proses bimbingan dan konseling secara lebih objektif dan terarah.

## PENDAHULUAN

Pemilihan jurusan pendidikan yang tepat di jenjang Sekolah Menengah Atas (SMA) maupun ke jenjang perguruan tinggi merupakan keputusan penting yang berdampak jangka panjang terhadap karier dan pengembangan diri siswa. Sayangnya, banyak siswa yang mengambil keputusan berdasarkan faktor eksternal seperti tekanan sosial atau tren sesaat, bukan berdasarkan pemahaman yang objektif terhadap minat, kemampuan, dan potensi diri mereka [9]

Di era digital dan informasi saat ini, kebutuhan terhadap sistem pendukung keputusan berbasis data semakin mendesak. Teknologi data mining memungkinkan proses analisis data siswa secara sistematis untuk menemukan pola yang tersembunyi dari hasil tes minat, nilai akademik, serta respons terhadap kuisioner psikometrik [3]. Salah satu metode klasifikasi yang efektif untuk permasalahan ini adalah algoritma C4.5, karena mampu menangani data dengan struktur hierarkis dan menghasilkan decision tree yang mudah diinterpretasi oleh pengguna awam [6].

Dengan mengembangkan sistem berbasis web, informasi hasil analisis dapat diakses secara fleksibel oleh siswa, guru BK, dan orang tua. Sistem ini tidak hanya mempermudah proses identifikasi jurusan yang tepat tetapi juga memberikan justifikasi logis berbasis pohon keputusan. Oleh karena itu, penelitian ini bertujuan membangun sistem rekomendasi jurusan pendidikan berbasis web yang mengintegrasikan algoritma C4.5 untuk menghasilkan klasifikasi yang akurat dan cepat.

## I. METODE

### 1.1. Data Mining

Data mining adalah proses menemukan pola tersembunyi, hubungan, dan informasi yang bermakna dari kumpulan data besar dengan menggunakan teknik statistik, pembelajaran mesin, dan kecerdasan buatan [13].

Menurut [14], data mining merupakan bagian dari proses KDD (Knowledge Discovery in Databases) yang berfokus pada ekstraksi pengetahuan yang sebelumnya tidak diketahui dan berguna dari data dalam jumlah besar.

#### Karakteristik Utama Data Mining:

- Bersifat otomatis atau semi-otomatis
- Menggunakan metode klasifikasi, klustering, asosiasi, dan deteksi anomali
- Berorientasi pada eksplorasi pola yang dapat dimanfaatkan untuk pengambilan keputusan

#### Fungsi-fungsi Utama Data Mining:

- Classification: pengelompokan data ke dalam kelas yang telah ditentukan
- Clustering: pengelompokan data berdasarkan kemiripan alami
- Association Rules: menemukan hubungan antar variabel dalam dataset
- Regression: prediksi nilai kontinu berdasarkan input data

### 1.2 Situs Web sebagai Media Sistem Informasi

Situs web adalah sekumpulan halaman digital yang saling terhubung dan dapat diakses melalui jaringan internet, yang disusun menggunakan protokol standar seperti HTTP/HTTPS dan bahasa markup seperti HTML [16].

Menurut [15], situs web dalam konteks sistem informasi berfungsi sebagai

antarmuka antara pengguna dan server data, mendukung interaksi seperti input, pengolahan, dan penyajian informasi secara real-time.

Situs Web Dinamis:

Berbeda dari situs statis, situs web dinamis mengintegrasikan pemrograman sisi server (seperti PHP, Python, Node.js) dan basis data (seperti MySQL, PostgreSQL) untuk menyediakan layanan interaktif dan personalisasi pengguna.

Peran dalam Sistem Rekomendasi:

Dalam penelitian ini, situs web digunakan sebagai media interaktif yang memungkinkan pengguna (siswa dan admin) untuk mengakses, menginput, dan memperoleh hasil rekomendasi secara real-time berbasis algoritma klasifikasi.

## 1.2. Metode C4.5

Algoritma adalah sekumpulan langkah sistematis yang terdefinisi secara jelas untuk menyelesaikan suatu masalah atau melakukan tugas komputasi tertentu. Algoritma harus memiliki awal dan akhir yang pasti, serta harus dapat diterapkan secara deterministik [17].

Menurut [18], algoritma adalah prosedur komputasi yang diuraikan secara baik dan terdiri dari serangkaian langkah instruksional yang, jika diikuti, akan menghasilkan output tertentu dari suatu input.

Pokok pemikiran dari perhitungan ini adalah mengejar pohon pilihan mengingat penentuan kualitas yang memiliki kebutuhan yang paling penting atau dapat disebut sebagai nilai tambah yang paling penting mengingat nilai entropi dari sifat tersebut sebagai pusat dari karakterisasi. Pada tahap ini metode penyusunan C4.5 memiliki dua standar kerja, yaitu: Mengejar pohon pilihan, dan membuat rencana keuangan (rule model). Rencana keuangan

memutuskan bahwa dibingkai oleh pohon pilihan akan membuat kondisi jika, struktur.

Ada empat tahapan dalam proses pemilihan tree decision dalam perhitungan C4.5, lebih spesifiknya:

a. Pilih properti untuk menjadi akarnya, berdasarkan nilai tambah terbesar dari atribut saat ini.

b. menentukan cabang untuk setiap nilai, adalah menentukan cabang terkoordinasi menggunakan nilai terbesar dari nilai kenaikan variable.

c. Partisi setiap kasus pada cabang, seperti yang ditunjukkan oleh perhitungan nilai tambahan terbesar dan estimasi dilakukan ketika estimasi nilai kenaikan tertinggi yang mendasari selesai dan kemudian proses estimasi kenaikan tertinggi selesai lagi tanpa menyertakan nilai variabel kenaikan yang mendasarinya.

d. Mengulangi siklus di setiap akar sehingga semua variable di akar memiliki kelas yang sama, dan diulangi seluruh proses pemrosesan yang paling penting untuk setiap bagian dari kasus hingga saat ini tidak mungkin diharapkan untuk menyelesaikan koneksi komputasi.

Klasifikasi C4.5 secara rekursif mengunjungi setiap pusat yang dipilih, memutuskan divisi terbaik, dengan tujuan tidak dibagi lagi. Dari tiga peneliti yang telah selesai, divisi terkoordinasi termasuk Komputasi C4.5 dikunci oleh para spesialis yang diperlukan untuk membantu mencari pilihan yang lebih mudah dan lebih cepat.

Ada beberapa cara dalam menentukan tree decision dengan perhitungan C4.5, khususnya:

1) Dapatkan informasi persiapan siap. Mempersiapkan informasi biasanya didasarkan pada informasi otentik yang telah terjadi dan telah dikelompokkan ke dalam kategori nya masing masing.

2) Tentukan fondasi pohon. akar akan diambil dari merek dagang yang dipilih dengan mengerjakan nilai Peningkatan dari

setiap properti, nilai Manfaat yang paling penting adalah akar dasar. Sebelum menegaskan nilai Ekspansi properti, pertama-tama hitung nilai entropi, untuk lebih jelasnya :

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

Keterangan :

- S : Himpunan Jumlah Kasus
- n : Jumlah Partisi S
- pi : Proporsi dari Si terhadap S

3) Kemudian hitung nilai Gain dengan metode information gain: Rumus Gain :

$$\text{Gain(S,A)=}$$

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

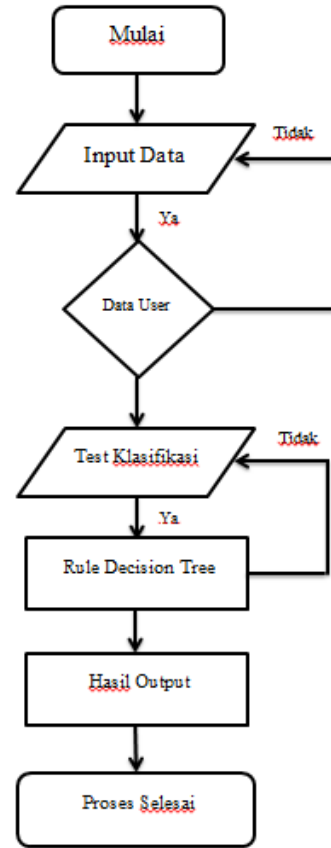
Keterangan :

- S : Himpunan Kasus
- A: Atribut
- n : Jumlah partisi atribut A
- |S<sub>i</sub>| : Jumlah kasus pada partisi ke-i
- |S| : jumlah kasus pada partisi S

## II. HASIL

[13] Klasifikasi adalah metode yang terlibat dengan menyelidiki informasi untuk mengungkap model yang menggambarkan kelas yang dikandungnya. Sedangkan model adalah classifier. Digunakan untuk mengakumulasi kelas-kelas yang terkandung dalam informasi. Misalnya, dalam Algoritma C4.5 sebuah kelas digambarkan sebagai pohon pilihan.

Berikut adalah proses klasifikasi pada aplikasi :



Gambar 1. Proses Klasifikasi Kepribadian

Pohon pilihan digunakan untuk menginterpretasikan nilai-nilai dalam pandangan satu segmen dari nilai-nilai dalam terang bagian lain. Cara kerja elaborasi pohon pilihan adalah dengan memanfaatkan tema persepsi atau memanfaatkan kemampuan untuk mengisolasi kelas informasi satu sama lain. Objek-objek istimewa dibagi-bagikan ke dalam beberapa klasifikasi perilaku dan sifat-sifat berdasarkan kumpulan-kumpulan yang bercirikan. Ketertiban mungkin digunakan ketika nilai objektif dapat diklasifikasikan. Dengan asumsi bahwa kualitasnya memadai, layak untuk membuat pohon pilihan yang mengelompokkan setiap item dalam set persiapan secara akurat.

## III. PEMBAHASAN

Pada tahap ini dilakukan pengujian terhadap sistem rekomendasi jurusan pendidikan berbasis algoritma C4.5 yang telah dikembangkan. Pembahasan mencakup implementasi sistem, pengujian fungsionalitas, analisis hasil klasifikasi, dan evaluasi akurasi sistem.

### 3.1 Implementasi Sistem

Sistem ini diimplementasikan sebagai aplikasi web menggunakan bahasa pemrograman PHP dengan basis data MySQL. Pengguna sistem terdiri atas dua pihak, yaitu admin dan siswa. Admin memiliki otoritas untuk mengelola data siswa, soal tes, dan hasil klasifikasi. Sementara itu, siswa dapat mengakses fitur tes minat, mengisi kuisioner, dan menerima hasil rekomendasi jurusan secara langsung melalui antarmuka pengguna web. Setiap jawaban dari kuisioner siswa menjadi atribut masukan bagi algoritma C4.5 untuk proses klasifikasi.

### 3.2 Proses Klasifikasi Menggunakan Algoritma C4.5

Algoritma C4.5 digunakan untuk membentuk pohon keputusan dari data pelatihan yang sudah diklasifikasikan sebelumnya. Atribut-atribut seperti nilai mata pelajaran tertentu (misal: Matematika, Bahasa Indonesia, IPA, IPS), hasil tes minat, dan preferensi siswa dijadikan variabel independen dalam membentuk model pohon. Pohon keputusan tersebut digunakan sebagai acuan sistem dalam memberikan rekomendasi jurusan kepada siswa, seperti jurusan IPA, IPS, atau Bahasa.

### 3.3 Pengujian Akurasi Sistem

Untuk mengetahui tingkat keakuratan sistem, dilakukan pengujian terhadap 50 data uji. Sistem dibandingkan dengan hasil klasifikasi manual yang dilakukan oleh guru bimbingan dan konseling. Dari hasil uji, sistem berhasil mengklasifikasikan 46 dari

50 data secara tepat, menghasilkan akurasi sebesar 92%. Perhitungan akurasi dilakukan dengan rumus:

$$\text{Akurasi} = \left( \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Data Uji}} \right) \times 100\%$$

$$\text{Akurasi} = (46 / 50) \times 100\% = 92\%$$

### 3.4 Tampilan Antarmuka Sistem

Antarmuka sistem dirancang secara sederhana agar mudah digunakan oleh siswa dan admin. Terdapat halaman login, halaman pengisian kuisioner, halaman hasil klasifikasi, serta visualisasi pohon keputusan yang dihasilkan. Sistem memberikan hasil rekomendasi jurusan secara langsung dan transparan dengan menyertakan alasan berdasarkan cabang pohon keputusan.

### 3.5 Analisis dan Interpretasi

Berdasarkan hasil pengujian dan observasi, sistem menunjukkan performa yang baik dalam memberikan rekomendasi jurusan. Pohon keputusan yang dihasilkan oleh algoritma C4.5 terbukti dapat digunakan sebagai dasar pengambilan keputusan karena memberikan jalur logis dari input ke output. Beberapa atribut yang paling menentukan dalam proses klasifikasi antara lain nilai Matematika, hasil tes logika, dan minat terhadap bidang ilmu tertentu.

Selain itu, sistem ini mampu menyesuaikan dengan data baru melalui proses pembelajaran tambahan. Artinya, pohon keputusan dapat diperbarui secara berkala seiring dengan bertambahnya data pelatihan, sehingga meningkatkan akurasi sistem di masa depan.

## IV. KESIMPULAN DAN SARAN

### 4.1. Kesimpulan

Sistem rekomendasi jurusan berbasis web yang dikembangkan dengan algoritma C4.5 terbukti efektif dan efisien dalam membantu proses bimbingan siswa. Sistem ini

memberikan hasil yang akurat dan dapat digunakan oleh siswa dan guru secara praktis.

#### **4.2. Saran**

Dalam penelitian ini aplikasi ini masih memiliki dan mempunyai banyak kekurangannya. supaya aplikasi ini bisa berfungsi lebih penulis memiliki beberapa saran, yaitu :

1. Perlu dilakukan tes dgn metode algoritma lain seperti apriori , k-means untuk membandingkan tingkat akurasi nya .
2. Dibuat lebih menarik dan lebih mudah sehingga lebih banyak orang yang memakai nya.

## DAFTAR PUSTAKA

- [1] B. Huda and B. Priyatna, "Penggunaan Aplikasi Content Manajement System (CMS) Untuk Pengembangan Bisnis Berbasis E-Commerce," *SYSTEMATICS*, vol. 1, no. 2, pp. 81–88, 2019.
- [2] Populix, "Apa itu Data Mining? Pengertian, Metode, Tahapan, dan Contoh Terbaru," Aug. 25, 2021.
- [3] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," *Jurnal Teknik Komputer*, vol. 4, no. 1, pp. 156–161, 2018.
- [4] M. Destiningrum and Q. J. Adrian, "SISTEM INFORMASI PENJADWALAN DOKTER BERBASIS WEB DENGAN MENGGUNAKAN FRAMEWORK CODEIGNITER," *Jurnal Teknoinfo*, vol. 11, no. 2, pp. 30–37, 2017.
- [5] G. G. Maulana, "PEMBELAJARAN DASAR ALGORITMA DAN PEMROGRAMAN MENGGUNAKAN EL-GORITMA BERBASIS WEB," *Jurnal Teknik Mesin (JTM)*, pp. 69–73, 2017.
- [6] R. Suqma, R. Halfis, and A. Hermawan, "Klasifikasi Channel Youtube Indonesia Menggunakan Algoritma C4.5," *Jurnal Teknik Informasi AMIS BSI*, pp. 21–28, 2019.
- [7] D. D. J. T. ;, ; Sitingjak, Maman, and J. Suwita, "ANALISA DAN PERANCANGAN SISTEM INFORMASI ADMINISTRASI KURSUS BAHASA INGGRIS PADA INTENSIVE ENGLISH COURSE DI CILEDUG TANGERANG," *JURNAL IPSIKOM*, vol. 8, no. 1, pp. 1–10, 2020.
- [8] Nawassyarif, M. Julkarnain, and K. R. Ananda, "SISTEM INFORMASI PENGOLAHAN DATA TERNAK UNIT PELAKSANA TEKNIS PRODUKSI DAN KESEHATAN HEWAN BERBASIS WEB," *Jurnal JINTEKS*, pp. 32–39, 2020.
- [9] B. A. Karim, "Teori Kepribadian dan Perbedaan Individu," *Education and Learning Journal*, vol. 1, no. 1, pp. 40–49, 2020.
- [10] Moh. P. H. Setiawan and F. Masya, "ANALISA PERANCANGAN SISTEM INFORMASI PELAPORAN FASILITAS UMUM DAN INFORMASI PEMBUATAN E-KTP UNTUK MASYARAKAT," *Jurnal Teknologi dan Sistem Informasi Univrab*, pp. 1–8, 2020.
- [11] O. Sari, A. Abdilah, and Sunarti, *WEB PROGRAMMING*. Yogyakarta: Graha ilmu, 2019.
- [12] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37–54.
- [13] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier/Morgan Kaufmann.
- [14] Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann.
- [15] Laudon, K. C., & Laudon, J. P. (2020). *Management Information Systems: Managing the Digital Firm* (16th ed.). Pearson Education.
- [16] Morville, P., & Rosenfeld, L. (2006). *Information Architecture for the World Wide Web* (3rd ed.). O'Reilly Media.
- [17] Weiss, M. A. (2012). *Data Structures and Algorithm Analysis in Java* (3rd ed.). Pearson Education.

- [18] Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms* (3rd ed.). MIT Press.

### **BIOGRAFI**

**Edy**, dosen tetap Program Studi Teknik Perangkat Lunak di Fakultas Sains dan Teknologi Universitas Buddhi Dharma.

**Rino**, dosen tetap Program Studi Teknik Informatika di Fakultas Sains dan Teknologi Universitas Buddhi Dharma.

**Desiyanna Lasut**, dosen tetap Program Studi Teknik Informatika di Fakultas Sains dan Teknologi Universitas Buddhi Dharma.