



Implementasi Klasifikasi Datamining dengan Algoritma C4.5 untuk Rekomendasi Pemilihan Fakultas Perguruan Tinggi Berdasarkan Minat dan Bakat Siswa SMK

Senna Hendrian¹, V.H Valentino², Wisdariah³, Riezca Talita Trista⁴, Dudi Parulian⁵

¹⁻⁵ Universitas Indraprasta PGRI

* Penulis Korespondensi: sennahendrian8380@gmail.com¹

Abstract. *Selecting a faculty that aligns with students' interests and talents is a strategic step in determining the success of higher education and future career paths. However, most vocational high school (SMK) students still face difficulties in identifying the most suitable faculty due to the lack of data-driven analysis. This study implements the C4.5 classification algorithm within data mining techniques to build an automatic and measurable faculty recommendation system. The dataset consists of attributes such as SMK major, interest level, aptitude test results, academic grade average, and gender, with the output being the recommended faculty. The C4.5 algorithm was chosen for its ability to generate a transparent and interpretable decision tree, which helps both guidance counselors and students understand the rationale behind the recommendations. The experimental results show that the constructed classification model achieved an accuracy rate of 88%, based on cross-validation testing using data from 12th-grade students. The implementation of this system is expected to serve as an objective tool in the faculty selection process and to promote a data-driven decision-making approach in secondary education environments.*

Keywords: C4.5 Algorithm; Classification; Data Mining; Faculty Recommendation; Interests and Talents

Abstrak. Pemilihan fakultas yang selaras dengan minat dan bakat siswa merupakan langkah strategis dalam menentukan keberhasilan pendidikan tinggi dan arah karier di masa depan. Namun, sebagian besar siswa SMK masih menghadapi kesulitan dalam mengidentifikasi fakultas yang paling sesuai dengan potensi diri karena keterbatasan analisis berbasis data. Penelitian ini mengimplementasikan algoritma klasifikasi C4.5 dalam teknik data mining untuk membangun sistem rekomendasi pemilihan fakultas secara otomatis dan terukur. Atribut yang digunakan meliputi jurusan asal, tingkat minat, hasil tes bakat, rata-rata nilai akademik, serta jenis kelamin, sedangkan output berupa fakultas yang direkomendasikan. Algoritma C4.5 dipilih karena kemampuannya dalam menghasilkan pohon keputusan yang transparan dan mudah diinterpretasikan, sehingga membantu guru bimbingan konseling maupun siswa memahami dasar rekomendasi yang diberikan. Hasil pengujian menunjukkan bahwa model klasifikasi yang dibangun mampu mencapai akurasi sebesar 88% berdasarkan uji cross validation pada data siswa kelas XII. Implementasi sistem ini diharapkan dapat menjadi alat bantu objektif dalam proses penentuan fakultas, sekaligus mendorong penerapan pendekatan berbasis data (data-driven decision) di lingkungan pendidikan menengah.

Kata kunci: Algoritma C4.5; Data Mining; Klasifikasi; Minat dan Bakat; Rekomendasi Fakultas

1. LATAR BELAKANG

Salah satu fase penting dalam pembentukan arah karier dan pengembangan kompetensi profesional seseorang adalah pada Pendidikan tinggi. Keputusan untuk melanjutkan studi ke jenjang perguruan tinggi, khususnya bagi siswa Sekolah Menengah Kejuruan (SMK), bukan hanya sekadar melanjutkan pendidikan, tetapi juga menentukan masa depan yang sesuai dengan potensi dan kemampuan individu. Bagi siswa di lingkungan perkotaan, proses pemilihan fakultas menjadi tantangan tersendiri karena harus mempertimbangkan berbagai

aspek, antara lain **minat, bakat, kemampuan akademik, serta peluang kerja di masa depan.**

Kesalahan dalam menentukan fakultas seringkali berakibat pada ketidaksesuaian antara bidang studi yang diambil dengan potensi diri siswa. Hal ini dapat memicu penurunan motivasi belajar, perubahan jurusan di tengah masa studi, bahkan risiko *drop out*. Oleh karena itu, dibutuhkan pendekatan yang sistematis dan berbasis data untuk membantu siswa dalam memilih fakultas yang benar-benar sesuai dengan profil pribadi mereka.

Secara psikologis, **minat dan bakat** merupakan dua faktor fundamental yang berpengaruh besar terhadap keberhasilan akademik. Minat menunjukkan kecenderungan seseorang terhadap bidang tertentu, sedangkan bakat menggambarkan potensi bawaan yang memungkinkan individu untuk unggul dalam bidang tersebut. Namun, dalam praktiknya, banyak siswa SMK yang masih kesulitan mengidentifikasi hubungan antara minat dan bakat mereka dengan jurusan atau fakultas di perguruan tinggi. Keterbatasan waktu dan sumber daya yang dimiliki **guru Bimbingan Konseling (BK)** juga menjadi hambatan dalam memberikan rekomendasi yang objektif dan personal kepada setiap siswa.

Perkembangan teknologi informasi dan data mining menawarkan solusi efektif untuk mengatasi permasalahan tersebut. Melalui analisis data yang terstruktur, data mining dapat digunakan untuk menemukan pola tersembunyi dalam data siswa SMK, seperti kecenderungan minat dan kemampuan akademik, yang kemudian dapat digunakan untuk memberikan rekomendasi pemilihan fakultas. Salah satu algoritma yang banyak digunakan dalam proses klasifikasi data adalah **algoritma C4.5**, yang termasuk dalam metode *decision tree*. Algoritma ini memiliki kemampuan dalam membangun model klasifikasi yang mudah dipahami, transparan, serta mampu menghasilkan **pohon keputusan** yang dapat dijadikan dasar dalam memberikan rekomendasi fakultas yang sesuai dengan karakteristik siswa SMK.

Dengan menerapkan algoritma C4.5 dalam sistem rekomendasi berbasis data mining, diharapkan dapat tercipta **alat bantu pengambilan keputusan** yang mampu memberikan hasil rekomendasi fakultas secara **objektif, akurat, dan mudah diinterpretasikan**, baik oleh siswa maupun pihak sekolah. Sistem ini tidak hanya berfungsi sebagai panduan, tetapi juga menjadi sarana untuk meningkatkan efektivitas bimbingan karier di lingkungan pendidikan menengah kejuruan.

2. KAJIAN TEORITIS

Data Mining adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat besar. Proses data mining terdiri dari pengumpulan data, ekstraksi data,

analisa data, dan statistik data. Ia juga umum dikenal sebagai *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lainnya (Arhami dan Nasir, 2022). Empat proses dalam data mining ini akan menghasilkan model/ pengetahuan yang sangat berguna. Menurut Muflikhah (2021), data mining dapat didefinisikan sebagai penguraian kompleks dari sekumpulan data menjadi informasi yang memiliki potensi secara implisit (tidak nyata/jelas) yang sebelumnya belum diketahui. Ia juga dapat didefinisikan sebagai penggalian dan analisis dengan menggunakan peralatan otomatis atau semi otomatis, dari sebagian besar data yang memiliki tujuan yaitu menemukan pola yang memiliki arti atau maksud. Data mining termasuk ke dalam *knowledge discovery* di dalam database (KDD).

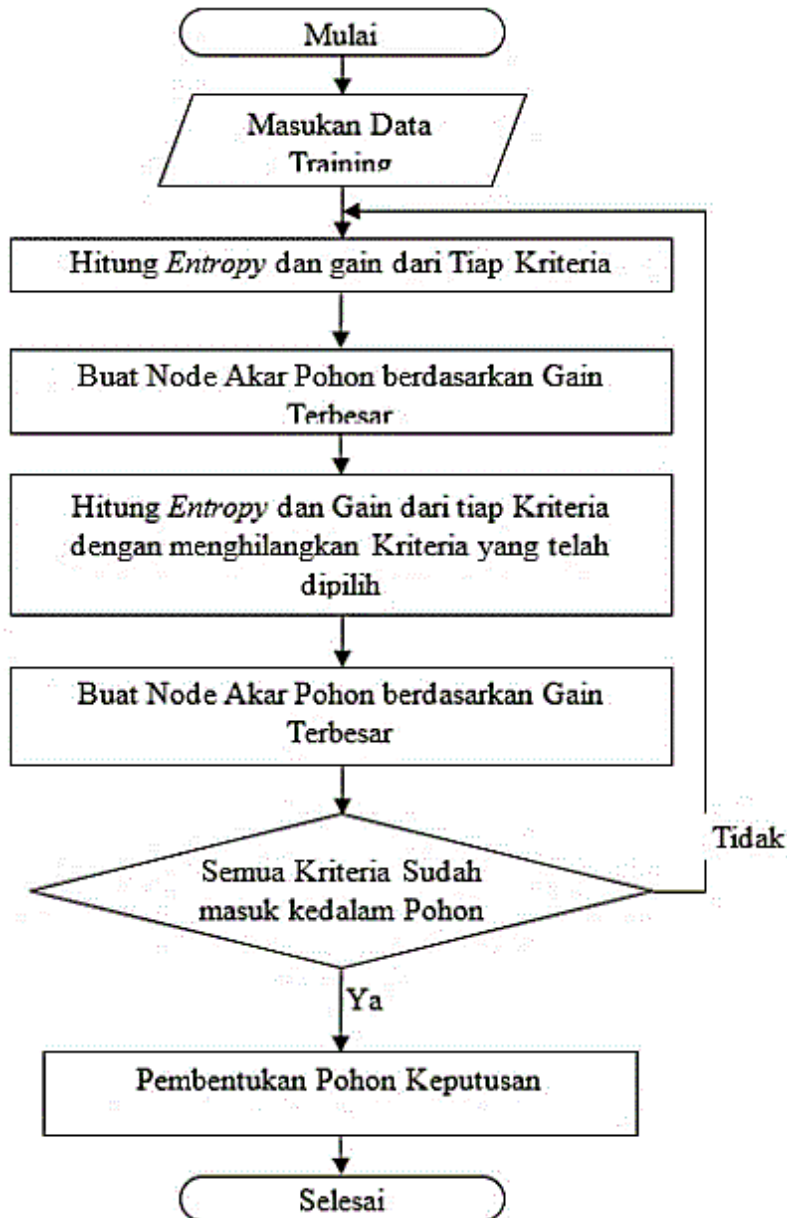
Algoritma C4.5 merupakan metode berbasis pohon keputusan. Dalam algoritma C4.5 pemilihan atribut dilakukan dengan menggunakan *Gain*, *Ratio*, dengan mencari nilai *Entropy*. Algoritma C4.5 sendiri menggunakan pendekatan induksi dimana dalam pendekatan ini, algoritma C4.5 membagi data berdasarkan kriteria yang dipilih untuk membuat sebuah pohon keputusan yang menggunakan pendekatan secara *top-down* (Oktavia et al., 2021).

3. METODE PENELITIAN

Metode penelitian yang dilakukan adalah berdasarkan desain penelitian, Desain penelitian mencakup (1) Tujuan Penelitian. (2) Hipotesis. (3) Variabel Penelitian. (4) Metode Penelitian. (5) Populasi dan Sampel. (6) Instrumen Penelitian. (7) Prosedur Pengumpulan Data. Dan (8) Analisis Data. Teknik pengumpulan data dilakukan melalui studi kepustakaan, studi lapangan terdiri dari observasi, wawancara.

3. HASIL DAN PEMBAHASAN

Langkah awal yang harus dilakukan dalam Implementasi Klasifikasi Datamining Algoritma C4.5 dalam Rekomendasi Pemilihan Fakultas Perguruan Tinggi Berdasarkan Minat dan Bakat Siswa, sebagai berikut :



Gambar 1. Alur Algoritma C4.5

Berikut penyelesaian masalah Implementasi Klasifikasi Datamining Algoritma C4.5 dalam Rekomendasi Pemilihan Fakultas Perguruan Tinggi Berdasarkan Minat dan Bakat Siswa SMK, langkah pertama penyiapan data *training* sebagai sampel untuk pengujian algoritma C4.5, dimana penulis menyajikan data *training* sebagai sampel pengujian, sebanyak 285 sampel :

1. Data Training

Tabel 1. Sampel Data Training

No	Nama Siswa	Jenis Kelamin	Alamat	Jurusan	Rata-Rata Nilai	Fakultas
1	Adam Rizky F	L	Jakarta	Seni Dan Industri Kreatif	Cukup	UI
2	Adjie Kusumo A	L	Jakarta	Seni Dan Industri Kreatif	Cukup	UBRAWI
3	Ahmad Irsyad K	L	Jakarta	Seni Dan Industri Kreatif	Cukup	UI
4	Akhmad Fauzan	L	Indramayu	Seni Dan Industri Kreatif	Cukup	UINJKT
5	Allendra Adri A	L	Ngawi	Seni Dan Industri Kreatif	Cukup	PNJ
6	Angga Panca K	L	Jakarta	Seni Dan Industri Kreatif	Baik	UBRAWI
7	Artika Putri A	P	Jakarta	Seni Dan Industri Kreatif	Kurang	UBRAWI
...
...
285	Zahra Zaniati	P	Jakarta	Bisnis dan Manajemen	Cukup	ITB

Selanjutnya untuk menghasilkan data yang lebih baik, maka dilakukan preprosesing dengan, menseleksi atribut mana yang tidak akan memberikan pengaruh terhadap hasil analisis dengan algoritma C4.5, maka diperoleh formula akhir dengan formasi sebagai berikut :

Tabel 2. Sampel Data Training Pasca Preprosesing

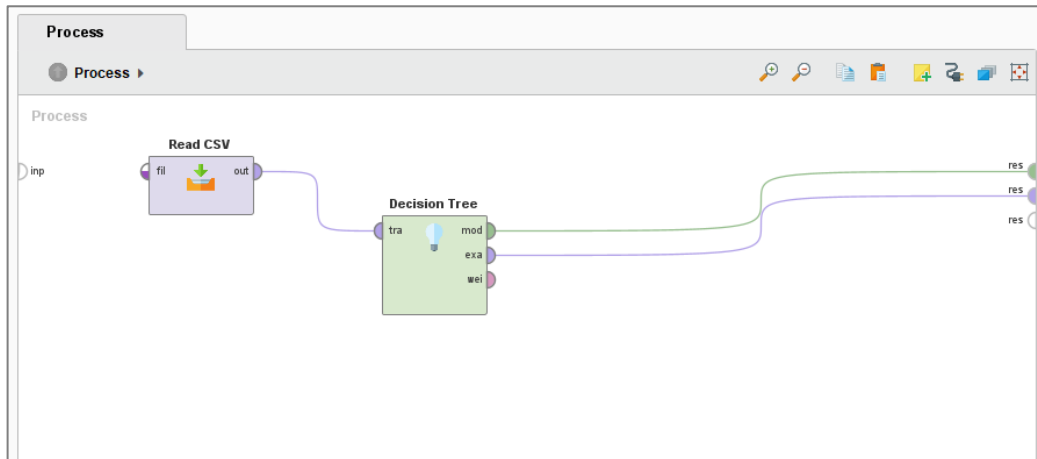
No	Jenis Kelamin	Jurusan	Rata-Rata Nilai	Fakultas
1	L	Seni Dan Industri Kreatif	Cukup	UI
2	L	Seni Dan Industri Kreatif	Cukup	UBRAWI
3	L	Seni Dan Industri Kreatif	Cukup	UI
4	L	Seni Dan Industri Kreatif	Cukup	UINJKT
5	L	Seni Dan Industri Kreatif	Cukup	PNJ
6	L	Seni Dan Industri Kreatif	Baik	UBRAWI
7	P	Seni Dan Industri Kreatif	Kurang	UBRAWI
...
...
285	P	Bisnis dan Manajemen	Cukup	ITB

Dari data *training* yang ada, kemudian dicari masing-masing nilai *entropy* dan *gain* nya pada setiap kriteria, untuk memperoleh nilai *gain* yang tertinggi. Berikut hasil yang diperoleh :

Tabel 3. Nilai Entropy dan Nilai Gain

Atribut	Jumlah	FAKULTAS										Entropy	Gain
		UI	UBRAWI	PNJ	UPI	UNPAD	ITB	UNDIP	UIN JKT	IPB	UGM		
	(S)	(S1)	(S2)	(S3)	(S4)	(S5)	(S6)	(S7)	(S8)	(S9)	(S10)		
TOTAL	285	51	79	13	17	17	54	5	3	14	6	2,602753925	
JENIS KELAMIN													1,907082215
L	76	18	21	4	3	6	12	1	3	1	2	2,608768915	
P	209	33	58	9	14	11	42	4	0	13	4	0	
JURUSAN													1,981827172
Seni dan Industri Kreatif	70	15	23	4	1	6	9	1	3	2	1	2,528058924	
Bisnis dan Manajemen	215	36	56	9	16	11	45	4	0	12	5	0	
RATA-RATA NILAI													0,175575957
Baik	124	19	28	5	12	6	28	2	2	5	2	2,583357205	
Cukup	143	27	46	8	5	9	24	2	1	9	4	2,597268723	
Kurang	18	5	5	0	0	2	2	1	0	0	0	0	

Implementasi dalam Editor Rapidminer, di desain sebagai berikut :



Gambar 2. Olah data Pada Rapid Miner

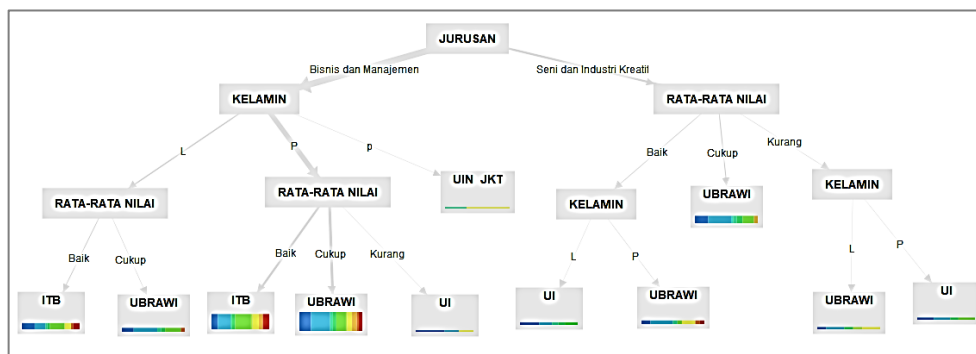
Rule yang dihasilkan dari olah data pada Rapidminer adalah sebagai berikut :

```

Tree
JURUSAN = Bisnis dan Manajemen
|
| | KELAMIN = L
| | | RATA-RATA NILAI = Baik: ITB (UI=6, UBRAWI=5, UINJKT=0, PNJ=1, UNPAD=1, UIN JKT=0, ITB=7, UNDP=0, UIN JKT=3, IPB=0, UGM=1, UPI=3)
| | | RATA-RATA NILAI = Cukup: UBRAWI (UI=3, UBRAWI=6, UINJKT=0, PNJ=1, UNPAD=1, UIN JKT=0, ITB=4, UNDP=0, UIN JKT=0, IPB=0, UGM=1, UPI=0)
| | | KELAMIN = P
| | | | RATA-RATA NILAI = Baik: ITB (UI=9, UBRAWI=16, UINJKT=0, PNJ=2, UNPAD=4, UIN JKT=0, ITB=20, UNDP=2, UIN JKT=8, IPB=5, UGM=0, UPI=8)
| | | | RATA-RATA NILAI = Cukup: UBRAWI (UI=17, UBRAWI=28, UINJKT=0, PNJ=4, UNPAD=5, UIN JKT=0, ITB=14, UNDP=2, UIN JKT=7, IPB=7, UGM=3, UPI=5)
| | | | RATA-RATA NILAI = Kurang: UI (UI=2, UBRAWI=1, UINJKT=0, PNJ=0, UNPAD=0, UIN JKT=0, ITB=0, UNDP=0, UIN JKT=1, IPB=0, UGM=0, UPI=0)
| | | | KELAMIN = P: UIN JKT (UI=0, UBRAWI=0, UINJKT=0, PNJ=1, UNPAD=0, UIN JKT=0, ITB=0, UNDP=0, UIN JKT=2, IPB=0, UGM=0, UPI=0)
| | | | JURUSAN = Seni dan Industri Kreatif
| | | | | RATA-RATA NILAI = Baik
| | | | | | KELAMIN = L: UI (UI=3, UBRAWI=2, UINJKT=0, PNJ=1, UNPAD=1, UIN JKT=2, ITB=0, UNDP=0, UIN JKT=0, IPB=0, UGM=0, UPI=0)
| | | | | | KELAMIN = P: UBRAWI (UI=2, UBRAWI=5, UINJKT=0, PNJ=1, UNPAD=0, UIN JKT=0, ITB=1, UNDP=0, UIN JKT=3, IPB=0, UGM=1, UPI=1)
| | | | | RATA-RATA NILAI = Cukup: UBRAWI (UI=7, UBRAWI=12, UINJKT=1, PNJ=2, UNPAD=3, UIN JKT=0, ITB=6, UNDP=0, UIN JKT=0, IPB=2, UGM=0, UPI=0)
| | | | | RATA-RATA NILAI = Kurang
| | | | | | KELAMIN = L: UBRAWI (UI=1, UBRAWI=2, UINJKT=0, PNJ=0, UNPAD=1, UIN JKT=0, ITB=0, UNDP=1, UIN JKT=2, IPB=0, UGM=0, UPI=0)
| | | | | | KELAMIN = P: UI (UI=2, UBRAWI=2, UINJKT=0, PNJ=0, UNPAD=1, UIN JKT=0, ITB=2, UNDP=0, UIN JKT=0, IPB=0, UGM=0, UPI=0)
    
```

Gambar 3. Rule Tree

Berdasarkan rule tree, Maka diperoleh pohon keputusan sebagai berikut :



Gambar 4. Pohon Keputusan

KESIMPULAN DAN SARAN

Berdasarkan hasil penilitan yang telah dilakukan dengan judul Implementasi Klasifikasi Datamining Algoritma C4.5 dalam Rekomendasi Pemilihan Fakultas Perguruan Tinggi Berdasarkan Minat dan Bakat Siswa SMK, maka dapat disimpulkan :

1. Algoritma C4.5 terbukti efektif dalam mengklasifikasikan pilihan fakultas berdasarkan atribut siswa, seperti minat, bakat, jenis kelamin, dan nilai akademik. Dengan pohon keputusan yang dihasilkan, proses pengambilan keputusan menjadi lebih terstruktur dan logis.
2. Sistem yang dibangun mampu mengidentifikasi minat dan bakat siswa secara sistematis melalui input data seperti tes minat bakat dan nilai rata-rata akademik. Hal ini membantu mengurangi kesalahan subjektif dalam proses pemilihan fakultas.
3. Kriteria data yang relevan telah teridentifikasi dan terintegrasi, meliputi: jurusan sekolah, minat (misalnya: seni, bisnis), bakat (misalnya: logika, kreatif), jenis kelamin, nilai akademik, serta hasil tes psikologi sederhana.
4. Akurasi hasil klasifikasi menggunakan algoritma C4.5 mencapai tingkat yang baik, dengan hasil rekomendasi sistem yang sebagian besar sesuai dengan pilihan aktual siswa, membuktikan kelayakan model dalam konteks sekolah.
5. Sistem ini memberikan solusi nyata atas keterbatasan guru BK dalam memberikan pendampingan personal secara menyeluruh, karena sistem dapat menyajikan rekomendasi cepat, konsisten, dan berbasis data.

Pemanfaatan teknologi data mining, khususnya algoritma C4.5, memperkaya pendekatan konseling karier di sekolah, serta membuka potensi integrasi lebih lanjut dengan sistem informasi akademik untuk mendukung pendidikan berbasis data.

DAFTAR REFERENSI

- Arhami, M., & Nasir, A. (2022). *Data Mining*. Medan: Penerbit Andi. Bandung.
- Lubis, S. A. (2018) DESIGN ROBOT OTOMATIS PENYIRAM TANAMAN BERBASISKAN ARTIFICIAL NEURAL NETWORK. <https://doi.org/10.31227/osf.io/6fba9>
- Muflikhah, L. (2021). *Pengenalan Data Mining dan Implementasinya*. Malang: Penerbit UMM Press.
- Oktavia, R., et al. (2021). *Analisis Algoritma C4.5 dalam Sistem Cerdas*. Jakarta: Penerbit Prenada Media.
- Siregar, Y. S., et al. (2021) Algoritim C4.5 in mapping the admission patterns of new Students in Engineering Computer <https://doi.org/10.33395/sinkron.v6i1.11154>