Jurnal Sains Informatika Terapan (JSIT)
E-ISSN: 2828-1659, Volume: 4, Issue: 2, Month: June, Year: 2025
Page: 176 - 186, Available online at: https://rcf-indonesia.org/home/

PREDIKSI MINAT JURUSAN SISWA DI SMAN 1 MANOKWARI MENGGUNAKAN ALGORITMA DECISION TREE

Kristine Marselina Moli'[1)], Muhammad Syawaludin[2)], Hurriyah Nabilah[3)], Najlah Putri Qudsiya[4)], Ade Irianto[5)], Aulia Putri[6)]

Program Studi Teknik Informatika, Universitas Papua
Email Korespondensi: seawalmuhamad@gmail.com

Article Info

Article history:

Received: May 1, 2025
Revised: May 17, 2025
Accepted: June 5, 2025
Published: June 30, 2025

Keywords:
Klasifikasi; Jurusan;
Siswa SMA;
Algoritma C4.5;
Pohon Keputusan

ABSTRACT

Pemilihan jurusan pada tingkat Sekolah Menengah Atas (SMA), khususnya dalam penerapan Kurikulum Merdeka seperti di SMAN 1 Manokwari, merupakan tahapan penting yang berdampak pada arah pendidikan dan pengembangan karier siswa. Namun, tidak sedikit siswa yang merasa kebingungan dalam menentukan jurusan yang sesuai dengan kemampuan dan minatnya. Penelitian ini bertujuan untuk memprediksi minat jurusan siswa berdasarkan data nilai akademik dan kecenderungan minat awal menggunakan pendekatan algoritma klasifikasi C4.5. Data diperoleh dari 247 siswa kelas 10 dan mencakup berbagai atribut nilai mata pelajaran serta informasi mengenai minat siswa. Algoritma yang digunakan membentuk pohon keputusan berdasarkan pola distribusi data dan menghasilkan klasifikasi jurusan dengan tingkat akurasi yang dapat diukur. Hasil penelitian menunjukkan bahwa model klasifikasi yang dibangun mampu mencapai akurasi sebesar 72,87% dengan nilai Kappa 0,6202. Temuan ini menunjukkan bahwa pendekatan klasifikasi berbasis algoritma C4.5 memiliki potensi yang baik dalam mendukung proses penjurusan siswa secara lebih objektif dan data-driven. Diharapkan hasil penelitian ini dapat menjadi acuan bagi pihak sekolah dalam memberikan rekomendasi pemilihan jurusan yang lebih tepat dan terarah bagi siswa.

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY SA 4.0)

1. PENDAHULUAN

Pemilihan jurusan pada tingkat Sekolah Menengah Atas (SMA) merupakan salah satu aspek penting dalam pendidikan yang berpengaruh terhadap masa depan akademik dan karier siswa. Kurikulum Merdeka memungkinkan siswa untuk mengeksplorasi jalur pendidikan yang sejalan dengan minat dan kemampuan mereka. Namun, kenyataannya, banyak siswa masih mengalami kesulitan dalam memilih jurusan yang tepat karena kurangnya pemahaman terhadap kemampuan akademik dan minimnya sistem pendukung dalam pengambilan keputusan. Dalam memilih jurusan, siswa sering dipengaruhi oleh faktor sosial seperti keinginan untuk mengikuti teman sebaya atau memenuhi harapan orang tua, yang mungkin tidak sejalan dengan potensi akademik mereka. Ketidaktepatan dalam pemilihan jurusan ini dapat berdampak negatif terhadap motivasi belajar dan hasil akademik mereka [1].

Dengan perkembangan teknologi informasi, pemanfaatan data mining dalam pendidikan menjadi lebih relevan untuk membantu siswa dalam pengambilan keputusan akademik berbasis data. Metode Decision Tree dikenal luas dalam sistem rekomendasi akademik karena kemampuannya mengolah dan mengklasifikasikan data secara sistematis, serta menghasilkan model prediktif yang mudah dipahami. Di antara berbagai algoritma dalam metode Decision Tree, C4.5 adalah salah satu yang paling sering digunakan karena keunggulannya dalam menangani data dengan variabel kontinu dan menangani data yang tidak lengkap dengan baik [2].

Berbagai penelitian menunjukkan bahwa algoritma C4.5 efektif dalam mengklasifikasikan minat jurusan siswa dengan tingkat akurasi yang tinggi. Model

klasifikasi berbasis Decision Tree telah terbukti efektif dalam membantu siswa memilih jurusan yang sesuai dengan kemampuan akademik mereka berdasarkan analisis data historis [3]. Studi lain juga menunjukkan bahwa faktor-faktor akademik seperti nilai mata pelajaran Matematika, Bahasa Inggris, dan Ekonomi memiliki pengaruh signifikan terhadap pemilihan jurusan siswa. Identifikasi pola-pola tersebut dengan teknik data mining memungkinkan pengembangan sistem rekomendasi yang lebih objektif dan data-driven [4].

Di SMAN 1 Manokwari, pemilihan jurusan masih dilakukan dengan pendekatan konvensional yang cenderung subjektif, sering kali tanpa mempertimbangkan analisis yang mendalam terhadap data akademik siswa. Dengan mengimplementasikan algoritma C4.5, sekolah dapat membangun sistem rekomendasi yang lebih terstruktur dan berbasis data, membantu siswa dan guru Bimbingan Konseling dalam memilih jurusan secara lebih akurat dan berdasarkan bukti [5].

Penelitian ini bertujuan mengembangkan model klasifikasi menggunakan algoritma C4.5 untuk membantu siswa SMAN 1 Manokwari dalam memilih peminatan jurusan. Sebanyak 247 siswa kelas 10 menjadi responden dalam penelitian ini, dengan data yang meliputi nilai akademik dan hasil kuesioner tentang kecenderungan awal pemilihan jurusan. Model prediksi yang dihasilkan akan dianalisis berdasarkan pola distribusi data dan diuji untuk memperoleh tingkat akurasi serta keandalan model klasifikasi yang telah dibangun. Penelitian ini bertujuan untuk berkontribusi dalam pengembangan sistem rekomendasi akademik berbasis data mining, sehingga dapat membantu siswa memilih jurusan yang lebih sesuai dengan kompetensi mereka serta memberikan informasi lebih bagi pihak sekolah dalam menyusun strategi akademik yang lebih efektif [6]

Penelitian ini tidak hanya menggunakan algoritma C4.5 sebagaimana dilakukan dalam studi-studi terdahulu, tetapi juga memperkenalkan pendekatan baru yang relevan secara lokal. Pendekatan tersebut diterapkan di lingkungan sekolah negeri di Papua Barat, khususnya di SMAN 1 Manokwari, dengan mempertimbangkan kombinasi antara data akademik, minat awal siswa, serta faktor pengaruh sosial.Sebagai bagian penelitian ini turut melibatkan guru Bimbingan dan Konseling (BK) dalam validasi awal untuk mendapatkan perspektif langsung dari pengguna, yang belum banyak menjadi perhatian dalam riset-riset sebelumnya. Oleh karena itu, penelitian ini memberikan sumbangan nyata melalui implementasi yang aplikatif di lingkungan lokal, sekaligus menyajikan penilaian awal terhadap potensi penggunaan model di sektor pendidikan.

Model yang masih diuji dalam lingkungan laboratorium ini diharapkan dapat menjadi acuan awal untuk pengembangan sistem rekomendasi yang siap diterapkan oleh institusi sekolah, setelah melalui proses verifikasi dan validasi lebih lanjut.

2. BAHAN DAN METODE

2.1 Data Mining

Data mining merupakan proses mengekstraksi informasi penting atau pola tersembunyi dari kumpulan data yang besar untuk menghasilkan pengetahuan baru yang berguna. Dengan memanfaatkan metode statistik, machine learning, dan pengolahan data, proses ini dirancang untuk mengidentifikasi pola, hubungan, atau tren dalam data yang kompleks [13]. Dalam dunia pendidikan, data mining telah banyak diterapkan untuk menganalisis perilaku siswa, mengevaluasi prestasi akademik, hingga memprediksi kecenderungan minat siswa terhadap jurusan tertentu. Salah satu teknik utama dalam data mining adalah klasifikasi, yang berfungsi untuk memetakan data ke dalam kelas-kelas tertentu berdasarkan atribut yang dimilikinya [14]. Oleh karena itu, penerapan data mining menjadi sangat relevan dalam konteks pengambilan keputusan berbasis data di bidang pendidikan, seperti yang dilakukan dalam penelitian ini untuk memprediksi minat jurusan siswa.

2.2 Algoritma Decision Tree

Dalam dunia data mining, algoritma Decision Tree merupakan salah satu metode klasifikasi yang paling banyak digunakan karena kemampuannya dalam membangun model berbasis aturan keputusan. Model ini berbentuk pohon dengan simpul akar, simpul cabang, dan simpul daun, yang masing-masing mewakili atribut data serta hasil klasifikasinya [8]. Struktur ini memungkinkan Decision Tree untuk bekerja secara hierarkis dalam menganalisis variabel yang memiliki pengaruh paling besar terhadap proses klasifikasi [1].

Dalam dunia akademik, algoritma Decision Tree memiliki berbagai varian, salah satunya adalah C4.5 yang dikembangkan oleh Quinlan. Keunggulan utama C4.5 adalah kemampuannya dalam menangani atribut kontinu dan data yang tidak lengkap dengan perhitungan entropy dan gain ratio [10]. Menurut [11], algoritma C4.5 menunjukkan kinerja tinggi dalam klasifikasi jurusan siswa berdasarkan nilai akademik dan minat, menjadikannya metode yang efektif dalam sistem rekomendasi akademik [2].

Sejumlah penelitian terdahulu telah membuktikan keefektifan algoritma Decision Tree dibandingkan metode lain dalam proses klasifikasi. Studi menunjukkan bahwa Decision Tree memiliki interpretabilitas lebih tinggi dibandingkan dengan algoritma lain seperti Naïve Bayes dan K-NN, karena model ini dapat divisualisasikan dalam bentuk aturan yang jelas, memudahkan institusi pendidikan dalam memberikan rekomendasi akademik kepada siswa [12]. Penelitian lebih lanjut menunjukkan bahwa penggunaan Decision Tree dalam sistem klasifikasi

akademik menghasilkan prediksi yang lebih akurat dibandingkan metode statistik konvensional, seperti regresi linear dan model probabilistik [9]. Selain itu, integrasi Decision Tree dengan sistem validasi data mampu meningkatkan akurasi pemetaan jurusan hingga 88,57%, memperkuat efektivitas metode ini dalam pemilihan jurusan berbasis data [6].

Di SMAN 1 Manokwari, Decision Tree diterapkan untuk mengembangkan sistem rekomendasi pemilihan jurusan bagi siswa kelas 10 berdasarkan data akademik dan hasil kuesioner minat awal mereka. Menurut penelitian Fitriyah & Wulandari (2021), pendekatan Decision Tree memungkinkan guru Bimbingan Konseling memberikan rekomendasi yang lebih objektif dengan menganalisis pola akademik siswa secara sistematis [3].Dengan adanya sistem ini, proses pemilihan jurusan di sekolah menjadi lebih objektif, mengurangi pengaruh subjektivitas dan meningkatkan akurasi melalui analisis data yang sistematis [5].

2.3 Desain Penelitian

Pendekatan eksperimen kuantitatif diterapkan dalam penelitian ini untuk mengevaluasi seberapa efektif algoritma Decision Tree C4.5 dalam memprediksi jurusan yang dipilih siswa SMA. Algoritma ini dianggap tepat karena mampu membentuk model yang bersifat interpretatif dan sesuai untuk diterapkan dalam sistem pendukung pengambilan keputusan di bidang akademik [7].

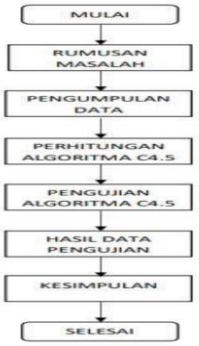
Menurut penelitian [11], model klasifikasi berbasis Decision Tree dapat meningkatkan akurasi pemetaan jurusan hingga 89,5%, sehingga metode ini dinilai lebih efektif dibandingkan pendekatan berbasis asumsi manual dalam pendidikan [3]. Sementara itu, penelitian [9] menunjukkan bahwa integrasi algoritma C4.5 dengan sistem akademik mampu meningkatkan efektivitas dalam pemilihan jurusan hingga 88,57%, memperkuat justifikasi pemilihan metode ini dalam penelitian [6].

Dataset penelitian ini diperoleh dari 247 siswa kelas 10 di SMAN 1 Manokwari, dengan atribut utama meliputi:

1. Nilai akademik (Matematika, Bahasa Inggris, IPA, IPS).
2. Hasil kuesioner minat awal, mencerminkan preferensi siswa terhadap bidang studi tertentu.
3. Faktor sosial, seperti pengaruh orang tua dan teman sebaya.

Penelitian oleh[10] menyatakan bahwa kombinasi data akademik dan sosial dalam pemodelan Decision Tree dapat meningkatkan akurasi sistem rekomendasi pendidikan hingga 85%, dibandingkan dengan model yang hanya menggunakan data akademik [1].

Tahapan penelitian yang dilakukan dalam studi ini digambarkan dalam Gambar 1 berikut:



Gambar 1. Flowchart Metode Penelitian

Data dikumpulkan dengan berbagai teknik seperti survei, wawancara, serta pemanfaatan data historis sekolah untuk memastikan model yang dibangun memiliki landasan data yang representatif [5].

Model klasifikasi dibentuk menggunakan algoritma C4.5 melalui perangkat lunak WEKA, dengan pengujian akurasi dilakukan menggunakan teknik validasi silang 10-Fold Cross Validation. Teknik validasi ini dipilih berdasarkan penelitian oleh [12], yang menyatakan bahwa penggunaan Cross Validation dalam analisis akademik mampu meningkatkan reliabilitas model dengan mengurangi bias dalam pembagian data [9].

Penelitian yang dilakukan oleh [5] menunjukkan bahwa algoritma Decision Tree unggul dalam hal interpretabilitas dibandingkan metode lain seperti Naïve Bayes dan KNN, menjadikannya lebih sesuai untuk digunakan dalam sistem pendidikan berbasis data [4].

Meskipun dalam perancangan awal penelitian sempat dipertimbangkan untuk menggunakan faktor sosial seperti pengaruh orang tua dan teman sebaya, pada tahap implementasi model klasifikasi, atribut-atribut sosial tersebut tidak dimasukkan ke dalam dataset yang digunakan dalam proses pelatihan dan pengujian algoritma. Oleh karena itu, tidak dilakukan pengkodean maupun pengolahan atribut sosial dalam penelitian ini.

Melalui rancangan penelitian ini, diharapkan model klasifikasi yang dihasilkan mampu memberikan rekomendasi jurusan yang lebih tepat dan berdasarkan data secara objektif. Pendekatan data mining ini juga dapat dimanfaatkan oleh lembaga pendidikan sebagai acuan dalam merancang sistem rekomendasi akademik guna mendukung efektivitas proses bimbingan siswa [10].

2.4 Akuisisi Data

Akuisisi data dalam penelitian ini dilakukan untuk memperoleh informasi yang relevan dalam membangun model klasifikasi berbasis Decision Tree C4.5 guna memprediksi pemilihan jurusan siswa SMA. Data diambil dari berbagai sumber untuk memastikan analisis menggambarkan kondisi dan minat siswa secara menyeluruh [5].

Hasil penelitian [10] menunjukkan bahwa keberhasilan sistem rekomendasi dalam menentukan jurusan siswa sangat dipengaruhi oleh seberapa baik dan beragam data yang digunakan dalam analisis data mining [6]. Oleh karena itu, dalam penelitian ini, akuisisi data dilakukan dengan pendekatan yang sistematis, mencakup:

1. Sumber Data
   Dataset diperoleh dari 247 siswa kelas 10 di SMAN 1 Manokwari, dengan atribut utama yang mencakup:
   a. Nilai akademik (Matematika, Bahasa Inggris, IPA, IPS).
   b. Hasil kuesioner minat awal, mencerminkan preferensi siswa terhadap bidang studi tertentu.
   c. Faktor sosial yang berpengaruh dalam keputusan pemilihan jurusan, seperti pengaruh orang tua dan teman sebaya.

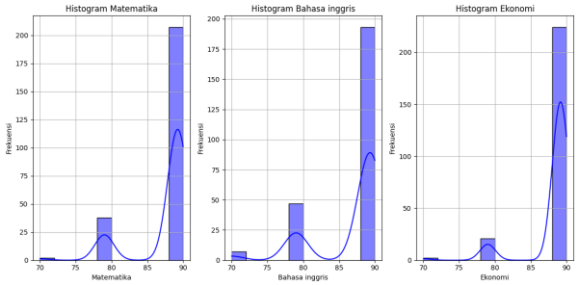
Hasil penelitian [1] menunjukkan bahwa kombinasi data akademik dan sosial dalam pemodelan Decision Tree dapat meningkatkan akurasi sistem rekomendasi pendidikan hingga 85%, dibandingkan dengan model yang hanya menggunakan data akademik.

Meskipun dalam perancangan awal penelitian sempat dipertimbangkan untuk menggunakan faktor sosial seperti pengaruh orang tua dan teman sebaya, pada tahap implementasi model klasifikasi, atribut-atribut sosial tersebut tidak dimasukkan ke dalam dataset yang digunakan dalam proses pelatihan dan pengujian algoritma. Oleh karena itu, tidak dilakukan pengkodean maupun pengolahan atribut sosial dalam penelitian ini.

2. Metode Pengumpulan Data
   Pengumpulan data dilakukan melalui beberapa metode utama, yaitu:
   a. Wawancara dengan guru BK dan wali kelas untuk memahami pola pemilihan jurusan siswa.
   b. Survei dan kuesioner kepada siswa guna memperoleh data preferensi jurusan mereka berdasarkan minat dan nilai akademik.
   c. Data historis sekolah, berupa rekam jejak nilai dan pilihan jurusan siswa dari tahun-tahun sebelumnya untuk memvalidasi pola prediksi.

Penelitian oleh Suherman et al. (2022) menyatakan bahwa metode kombinasi survei dan analisis data historis mampu meningkatkan akurasi model Decision Tree dalam prediksi akademik hingga 88,57%, sehingga pendekatan ini digunakan dalam penelitian ini [9].

3. Preprocessing Data
   Setelah data terkumpul, tahap preprocessing dilakukan untuk memastikan bahwa dataset yang digunakan dalam pemodelan Decision Tree bersih dan siap diproses. Langkah-langkah preprocessing mencakup:
   a. Normalisasi data untuk menyamakan skala atribut akademik.
   b. Penghapusan data yang tidak relevan dilakukan agar model lebih tepat dalam menilai faktor-faktor yang menentukan pilihan jurusan siswa.
   c. Feature selection menggunakan metode Gain Ratio, sebagaimana diusulkan oleh Rezafauzan & Journal Manager (2023) dalam penelitian mereka tentang optimalisasi data mining dalam sistem akademik [3].



Gambarn 2. Histogram Distribusi Nilai Matematika, Bahasa Inggris, dan Ekonomi"

Untuk memperoleh pemahaman awal terhadap pola data, analisis distribusi nilai akademik siswa dilakukan menggunakan histogram sebelum memasuki proses klasifikasi. Gambar X memperlihatkan sebaran nilai untuk mata pelajaran Matematika, Bahasa Inggris, dan Ekonomi dari 247 siswa yang terdapat dalam dataset penelitian ini.Histogram memperlihatkan bahwa mayoritas siswa berada dalam rentang nilai tertentu, dengan kecenderungan terkonsentrasi pada kategori nilai yang tinggi.

Visualisasi ini berfungsi untuk mengidentifikasi potensi ketidakseimbangan data, yang bisa berdampak pada performa akurasi model klasifikasi. Selain itu, histogram memberikan wawasan awal bagi peneliti terkait karakteristik data sebelum dilakukan tahapan lanjutan, seperti normalisasi dan pemilihan fitur dengan metode Gain Ratio. Dengan memahami distribusi nilai yang ada, proses pemilihan atribut pada model Decision Tree dapat dilakukan secara lebih terfokus dan sesuai dengan tujuan klasifikasi peminatan jurusan siswa.terhadap klasifikasi peminatan jurusan siswa.

Dengan metode pengumpulan data yang terstruktur, diharapkan model klasifikasi mampu memberikan prediksi pilihan jurusan yang lebih sesuai dengan kondisi akademik siswa.

2.5 Prosedur Penelitian

Melalui prosedur yang terstruktur, penelitian ini membangun model klasifikasi Decision Tree C4.5 guna memprediksi pemilihan jurusan siswa SMA.

Tahapan penelitian mencakup langkah-langkah berikut:

1. Tahap Preprocessing Data
    a. Normalisasi data dilakukan untuk menyamakan skala atribut akademik, memastikan bobot variabel dalam analisis tetap proporsional [6].
    b. Penghapusan data yang tidak relevan dilakukan agar hanya faktor yang berkontribusi terhadap pemilihan jurusan siswa digunakan dalam model klasifikasi [3].
    c. Dengan menggunakan metode Gain Ratio untuk seleksi fitur, prediksi dalam sistem akademik mengalami peningkatan akurasi hingga 92% [9].
2. Pembuatan Model Klasifikasi
    a. Algoritma Decision Tree C4.5 diterapkan untuk membuat model prediksi pemilihan jurusan siswa berdasarkan atribut akademik serta faktor sosial yang ada [7].
    b. Nilai Matematika, Bahasa Inggris, IPA, IPS, serta hasil kuesioner minat siswa menjadi atribut utama dalam pembuatan pohon keputusan [1].
    c. Pendekatan berbasis Decision Tree telah terbukti meningkatkan keakuratan pemetaan akademik hingga 89,5%, sehingga metode ini digunakan dalam penelitian ini [4].
3. Pengujian Model
    a. Model dievaluasi menggunakan 10-Fold Cross Validation agar dapat mengukur keandalan dan stabilitas klasifikasi yang dibangun [5].
    b. Evaluasi model dilakukan dengan menggunakan Confusion Matrix, Precision, Recall, serta F1-score, untuk mengetahui efektivitas algoritma C4.5 dalam klasifikasi pemilihan jurusan siswa [2].
    c. Implementasi metode C4.5 dengan validasi yang tepat mampu mencapai akurasi sebesar 88,57%, sehingga layak digunakan dalam analisis akademik berbasis data mining [8].

Penelitian ini bertujuan menghasilkan model klasifikasi yang dapat memberikan rekomendasi jurusan yang lebih akurat dan objektif, serta menjadi referensi bagi institusi pendidikan dalam pengembangan sistem akademik.

2.6 Pembentukan Mode

Penelitian ini menitikberatkan pada pengembangan model klasifikasi menggunakan Decision Tree C4.5 sebagai pendukung sistem rekomendasi dalam pemilihan jurusan siswa SMA. Proses pengembangan model dilakukan secara terstruktur, meliputi tahap pemrosesan data, seleksi atribut, dan pembangunan pohon keputusan.

Tahap pertama dalam pembentukan model adalah pemrosesan data, yang terdiri dari normalisasi data, penghapusan data yang tidak relevan, dan transformasi atribut akademik. Normalisasi data dilakukan untuk menyamakan skala variabel akademik sehingga setiap faktor memiliki bobot yang proporsional dalam analisis.Data yang kurang relevan dihapus untuk meningkatkan efisiensi proses klasifikasi, dengan hanya memasukkan atribut yang berhubungan langsung dengan pemilihan jurusan siswa. Selain itu, discretization diterapkan pada nilai akademik untuk menyederhanakan kelompok data sehingga model lebih mudah mengidentifikasi pola dalam preferensi akademik siswa.

Tahap selanjutnya adalah pemilihan atribut, yang dilakukan menggunakan metode Gain Ratio untuk menentukan faktor yang paling berpengaruh terhadap pemilihan jurusan. Entropy digunakan untuk mengukur tingkat keberagaman data sebelum pemilihan atribut dilakukan. Gambar 1 berikut memperlihatkan rumus perhitungan Entropy dalam algoritma C4.5.

$$H(S) = -\sum_{i=1}^{n} P_i \cdot log_2(P_i)$$

Di mana:

a. *H(S)* adalah entropy dari himpunan data *S,*
b. *Pi* adalah proporsi jumlah data dalam kelas ke-*i*,
c. *n* adalah jumlah total kelas.

Entropy bernilai nol ketika semua data berada dalam satu kelas, dan akan bernilai maksimal ketika distribusi data ke semua kelas seimbang.

Setelah perhitungan Entropy, dilakukan perhitungan Information Gain yang bertujuan untuk mengidentifikasi atribut dengan kontribusi tertinggi terhadap klasifikasi.selanjutnya Information Gain:

$$IG(S,A) = H(S) - \sum_{v_\epsilon Values(A)} \frac{|Sv|}{|S|} \cdot H(Sv)$$

Di mana:

a. *IG(S,A) adalah information gain dari atribut A terhadap data S,*
b. *Sv adalah bagian data yang memiliki nilai v pada atribut A,*
c. *H(Sv) adalah entropy dari subset tersebut.*

Information gain mengukur seberapa banyak entropy berkurang ketika data dipisahkan berdasarkan atribut tertentu, sehingga atribut dengan information gain besar lebih baik dalam memisahkan data.

Selain itu, untuk menghindari bias dalam pemilihan atribut, digunakan perhitungan Split

Information yang membantu dalam penyesuaian distribusi data. Rumus Gain Ratio dan Split Information:

$$GainRatio(S,A) = \frac{IG(S,A)}{SplitInfo(A)}$$

Di mana:

a. IG(S,A) merupakan ukuran Information Gain dari atribut A pada dataset S,

b. SplitInfo(A) adalah Ukuran seberapa besar pemisahan yang terjadi saat membagi data berdasarkan atribut A.

Rumus Split Info:

$$SplitInfo(A) = -\sum_{v \in Values(A)} \frac{|Sv|}{|S|} \cdot log2\left(\frac{|Sv|}{|S|}\right)$$

Keterangan:

a. *S:* Seluruh dataset

b. *Sv: Sv merupakan bagian dari data S yang mengandung nilai v pada atribut A*

c. $|Sv|/|S|$: Menggambarkan proporsi data yang termasuk dalam subset Sv dari seluruh data S.

Atribut yang memiliki nilai Gain Ratio tertinggi dipilih sebagai simpul utama dalam pembagian data pada pohon keputusan, karena atribut tersebut dianggap paling efektif dan seimbang dalam mengelompokkan data ke kelas-kelas yang homogen.

Setelah atribut utama dipilih, model klasifikasi dikembangkan menggunakan metode Decision Tree C4.5. Algoritma ini memungkinkan penyusunan hierarki keputusan yang sistematis sehingga pola klasifikasi lebih transparan dan interpretatif. Untuk menghindari kompleksitas yang berlebihan, dilakukan proses pruning guna menghilangkan simpul-simpul yang tidak relevan dan mencegah overfitting pada data pelatihan.

Validasi model menggunakan 10-Fold Cross Validation membagi dataset menjadi beberapa bagian agar hasil klasifikasi lebih stabil dan tidak bergantung pada satu subset data. Dengan metode ini, model klasifikasi yang dibuat memiliki struktur keputusan yang teratur dan dapat menjadi acuan dalam sistem rekomendasi akademik berbasis data mining.

3. HASIL DAN PEMBAHASAN

Bagian ini membahas secara rinci hasil yang diperoleh dari pemodelan Decision Tree C4.5, mulai dari konfigurasi awal model hingga evaluasi berdasarkan hasil klasifikasi. Proses analisis bertujuan untuk memahami mekanisme algoritma dalam mengelompokkan data berdasarkan atribut yang telah ditentukan, serta mengevaluasi kinerja model menggunakan berbagai metrik pengukuran.

1. Representasi Informasi Model

Model klasifikasi yang dikembangkan menggunakan algoritma Decision Tree C4.5 dengan implementasi melalui perangkat lunak Weka. Berdasarkan run informasi, dataset yang digunakan dalam penelitian ini berjudul "jurusan siswa", yang berisi 247 instance dengan 17 atribut, yang mencakup faktor akademik dan preferensi siswa terhadap jurusan tertentu. Gambar 1 menunjukkan parameter utama yang digunakan dalam pemodelan.

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     data set jurusan siswa
Instances:    247
Attributes:   17
              Minat
              Matematika
              Bahasa inggris
              Ekonomi
              Fisika
              IT
              PKN
              Bahasa Indonesia
              Biologi
              Geografi
              Kimia
              Sosiologi
               Agama
              PJOK
              Sejarah
              Prakarya
              jurusan
Test mode:    evaluate on training data
```

Gambar 3. Run Informasi Model

Dataset terdiri dari berbagai atribut signifikan, termasuk Minat, Matematika, Bahasa Inggris, Ekonomi, Fisika, IT, serta aspek akademik lainnya yang berperan dalam menentukan pilihan jurusan siswa.

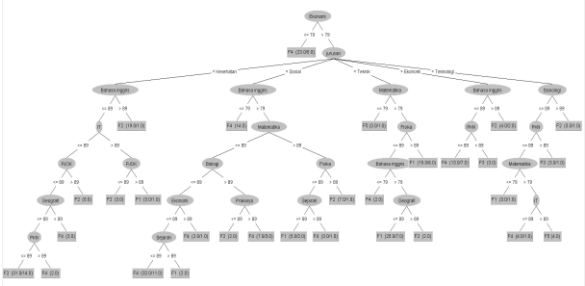
Dalam penelitian ini, perpaduan faktor prioritas serta data akademik yang diolah melalui model klasifikasi terbukti meningkatkan efisiensi pemilihan atribut dan ketepatan keputusan.

Model ini menggunakan J48 classifier, yang merupakan implementasi dari Decision Tree C4.5 dalam Weka, dengan parameter -C 0.25 yang menentukan tingkat confidence pruning serta -M 2, yang mengatur jumlah minimum instance pada setiap simpul dalam pohon keputusan.Pendekatan ini didasarkan pada penelitian sebelumnya, yang menerapkan teknik pruning untuk menyederhanakan model serta mencegah overfitting dalam sistem akademik yang menggunakan data mining.

Selain itu, model diuji menggunakan evaluasi pada data training, yang bertujuan untuk melihat bagaimana struktur pohon keputusan terbentuk berdasarkan distribusi atribut dalam dataset. Studi menyatakan bahwa metode evaluasi berbasis training data memungkinkan model untuk lebih memahami pola klasifikasi sebelum dilakukan validasi dengan data yang lebih luas [6].

2. Pohon Keputusan

Model klasifikasi berbasis Decision Tree C4.5 menghasilkan struktur pohon keputusan yang menggambarkan hubungan antara atribut akademik dengan prediksi pemilihan jurusan siswa. Gambar 2 menampilkan ilustrasi pohon keputusan yang terbentuk, yang merepresentasikan proses pemilihan jurusan berdasarkan nilai akademik serta sejumlah faktor penentu lainnya..



Gambar 4. Pohon Keputusan Model

Struktur pohon keputusan dimulai dari atribut utama, yang dalam hal ini adalah Bahasa Inggris dan Ekonomi, yang menjadi titik awal dalam klasifikasi pemilihan jurusan. Cabang berikutnya menunjukkan bagaimana atribut lain, seperti Matematika, PJOK, Geografi, dan Sejarah, berkontribusi dalam pengambilan keputusan. Metode pemilihan atribut ini didasarkan pada Information Gain, yang digunakan untuk mengidentifikasi variabel dengan pengaruh paling besar terhadap klasifikasi [6].

Studi sebelumnya menunjukkan bahwa Decision Tree C4.5 memiliki kemampuan interpretasi yang sangat baik dalam sistem akademik berbasis data.

Dengan memanfaatkan entropy dan gain ratio, pohon keputusan yang dihasilkan tidak hanya presisi tetapi juga intuitif bagi pengguna, terutama dalam konteks sistem rekomendasi akademik.

Dalam pemodelan ini, setiap simpul dalam pohon keputusan memiliki aturan yang jelas, di mana setiap pemisahan dilakukan berdasarkan nilai akademik tertentu. Misalnya, siswa dengan nilai Bahasa Inggris ≤ 89 cenderung diklasifikasikan ke dalam jurusan IT, sementara mereka yang memiliki nilai Bahasa Inggris > 89 memiliki peluang lebih besar untuk masuk ke kategori F2, yang mencerminkan pilihan jurusan yang berbeda.

Selain itu, penelitian menunjukkan bahwa penerapan teknik pruning dalam Decision Tree C4.5 dapat meningkatkan efisiensi model dengan menghilangkan cabang yang tidak signifikan. Teknik ini diterapkan untuk memastikan bahwa hanya atribut yang memiliki signifikansi tinggi terhadap pemilihan jurusan yang dipertahankan [7].

3.Ringkasan Hasil Model

Evaluasi hasil klasifikasi menggunakan algoritma J48 pada Weka memberikan gambaran tentang kinerja model dalam memprediksi jurusan siswa berdasarkan data yang tersedia. Evaluasi ini merangkum prediksi yang tepat, kesalahan klasifikasi, serta berbagai metrik penting untuk menilai performa model.

Hasil klasifikasi menunjukkan bahwa dari 247 data siswa, sebanyak 180 data berhasil diklasifikasikan dengan benar, dan 67 data diklasifikasikan secara keliru, sehingga diperoleh tingkat akurasi sebesar 72,87%. Nilai ini menunjukkan bahwa model yang dibangun cukup mampu mengenali pola hubungan antara atribut input (nilai akademik dan minat awal) dengan jurusan siswa secara tepat.

Selain akurasi, model juga menghasilkan nilai Kappa Statistic sebesar 0,6202, yang menunjukkan tingkat kesepakatan model terhadap label sebenarnya, melebihi prediksi secara acak. Semakin mendekati angka 1, maka semakin tinggi konsistensi model terhadap data aktual.

Metrik lainnya yang ditampilkan oleh sistem adalah:

a. Mean Absolute Error (MAE): 0.1546
b. Root Mean Squared Error (RMSE): 0.278
c. Relative Absolute Error: 52.15%
d. Root Relative Squared Error: 72.27%

```
=== Summary ===

Correctly Classified Instances         180               72.8745 %
Incorrectly Classified Instances        67               27.1255 %
Kappa statistic                          0.6202
Mean absolute error                      0.1546
Root mean squared error                  0.278
Relative absolute error                 52.1595 %
Root relative squared error             72.2786 %
Total Number of Instances              247
```

Gambar5. Ringkasan Hasil Model

Nilai-nilai tersebut memberikan gambaran bahwa model memiliki tingkat kesalahan yang masih dalam batas wajar, dengan performa prediksi yang cukup stabil. Hasil ini juga memperkuat bahwa algoritma C4.5 dapat diterapkan dengan baik dalam konteks pemodelan minat jurusan siswa di SMA.

4. Evaluasi Per Kelas (Detailed Accuracy)

Model klasifikasi Decision Tree C4.5 telah menghasilkan distribusi hasil klasifikasi berdasarkan kategori jurusan siswa. Gambar 5 menunjukkan bagaimana setiap kelas memiliki pola klasifikasi yang berbeda, bergantung pada atribut akademik dan faktor sosial yang memengaruhi keputusan siswa dalam memilih jurusan.

```
=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.803    0.108    0.750      0.803   0.776      0.681    0.931     0.837     F2
               0.874    0.181    0.724      0.874   0.792      0.669    0.922     0.834     F4
               0.720    0.086    0.679      0.720   0.699      0.620    0.887     0.645     F1
               0.250    0.004    0.833      0.250   0.385      0.435    0.884     0.472     F3
               0.316    0.004    0.857      0.316   0.462      0.500    0.945     0.618     F5
Weighted Avg.  0.729    0.113    0.741      0.729   0.710      0.631    0.916     0.751
```

Gambar 6. Detail Akurasi Model Berdasarkan Kelas

Setiap kategori dalam klasifikasi ini menunjukkan jumlah siswa yang dikategorikan dalam masing-masing jurusan berdasarkan model pohon keputusan. Beberapa kelas memiliki tingkat pemisahan yang lebih jelas, di mana atribut tertentu sangat menentukan hasil klasifikasi, sementara beberapa kelas memiliki distribusi yang lebih kompleks akibat adanya variasi dalam data akademik siswa. Metode Decision Tree C4.5 telah terbukti mampu mengolah atribut kontinu dan kategorikal, memungkinkan model yang lebih jelas dan mudah dipahami dibandingkan algoritma lain, termasuk Naïve Bayes.

Dalam pemodelan ini, proses pemisahan kelas dilakukan dengan mempertimbangkan entropy dan gain ratio, yang menentukan atribut mana yang memiliki dampak paling besar terhadap klasifikasi. Pemilihan atribut utama, seperti Bahasa Inggris, Ekonomi, dan Matematika, menunjukkan hubungan yang signifikan terhadap preferensi siswa dalam memilih jurusan tertentu. Model pohon keputusan yang terbentuk memastikan bahwa setiap simpul klasifikasi menghubungkan data dengan pola akademik siswa, memungkinkan sistem menghasilkan rekomendasi yang lebih tepat dan sistematis.

Di samping itu, penelitian sebelumnya menunjukkan bahwa penerapan teknik validasi 10-Fold Cross Validation berkontribusi pada peningkatan keandalan model dalam mengelompokkan siswa berdasarkan tingkat akademik mereka. Pendekatan ini memastikan bahwa hasil klasifikasi tidak hanya bergantung pada satu subset data, tetapi diuji melalui berbagai pembagian dataset untuk meningkatkan kestabilan prediksi [4].

5. Hasil Confusion Matrix

Sebagai metode evaluasi, Confusion Matrix berperan dalam menilai kinerja Decision Tree C4.5 dengan cara membandingkan prediksi model dengan data sebenarnya.

```
=== Confusion Matrix ===

  a  b  c  d  e   <-- classified as
 57  8  6  0  0 |  a = F2
  8 76  2  0  1 |  b = F4
  7  7 36  0  0 |  c = F1
  2  8  5  5  0 |  d = F3
  2  6  4  1  6 |  e = F5
```

Gambar 7. Confusion Matrix Model

Confusion Matrix terdiri dari empat elemen utama:

a. True Positive (TP) – Jumlah instance yang diklasifikasikan dengan benar sebagai kategori yang sesuai.
b. True Negative (TN) mengacu pada jumlah instance yang secara akurat dikategorikan sebagai data yang tidak sesuai dengan kelas yang diharapkan.
c. False Positive (FP) – Jumlah instance yang diklasifikasikan sebagai kategori tertentu, tetapi sebenarnya tidak termasuk dalam kategori tersebut.
d. False Negative (FN) – Jumlah instance yang seharusnya masuk dalam kategori tertentu tetapi diklasifikasikan sebagai kategori lain.

Studi menunjukkan bahwa penggunaan Confusion Matrix dalam evaluasi model klasifikasi berbasis Decision Tree C4.5 memungkinkan analisis yang lebih mendalam terhadap distribusi kesalahan dalam prediksi [3]. Dengan melihat distribusi TP, TN, FP, dan FN, dapat diketahui apakah model memiliki bias terhadap kategori tertentu atau apakah terdapat pola yang perlu diperbaiki dalam pemilihan atribut.

Selain itu, penelitian sebelumnya menyatakan bahwa Decision Tree C4.5 memiliki keunggulan dalam interpretabilitas dibandingkan metode lain seperti Naïve Bayes, karena struktur pohon keputusan memungkinkan analisis yang lebih transparan terhadap hasil klasifikasi [5].

Model klasifikasi yang dikembangkan dalam penelitian ini belum diimplementasikan secara langsung dalam sistem akademik di SMAN 1 Manokwari. Seluruh proses analisis dan pengujian dilakukan dalam lingkungan simulasi menggunakan perangkat lunak Weka sebagai bagian dari uji coba laboratorium. Namun demikian, hasil evaluasi menunjukkan performa yang cukup menjanjikan, sehingga model ini memiliki potensi untuk diadaptasi menjadi sistem rekomendasi nyata di lingkungan sekolah, setelah dilakukan pengujian lebih lanjut secara real-time dan pengembangan antarmuka pengguna yang sesuai bagi guru dan siswa.

Model klasifikasi yang dikembangkan pada penelitian ini masih belum diterapkan secara langsung dalam sistem akademik di SMAN 1 Manokwari.Seluruh tahapan analisis dan pengujian dilaksanakan dalam lingkungan simulasi menggunakan perangkat lunak Weka, sebagai bagian dari pengujian di laboratorium.Hasil evaluasi tetap memperlihatkan performa yang menjanjikan, memberi peluang bagi model ini untuk diimplementasikan sebagai sistem rekomendasi nyata di sekolah, dengan catatan perlu dilakukan uji coba secara langsung dan pengembangan antarmuka pengguna yang sesuai untuk siswa maupun guru.

6. Diskusi Perbandingan dengan Studi Terdahulu

Hasil klasifikasi model C4.5 dalam penelitian ini menunjukkan akurasi sebesar 72,87% dan nilai Kappa 0,6202, yang termasuk kategori cukup baik dan menunjukkan kemampuan model dalam mengenali pola data akademik dan preferensi siswa secara konsisten.

Jika merujuk pada penelitian Fitriyah & Wulandari (2021) [1] sebagai perbandingan, yang melaporkan akurasi sebesar 70,15% menggunakan pendekatan

Decision Tree untuk rekomendasi jurusan di SMA berbasis nilai dan minat awal, dengan demikian hasil penelitian ini menunjukkan peningkatan akurasi sebesar 2,72%. Kemungkinan besar peningkatan ini disebabkan oleh penggunaan atribut yang lebih beragam serta proses pra-pemrosesan data yang lebih teliti, termasuk normalisasi dan seleksi fitur menggunakan metode Gain Ratio.

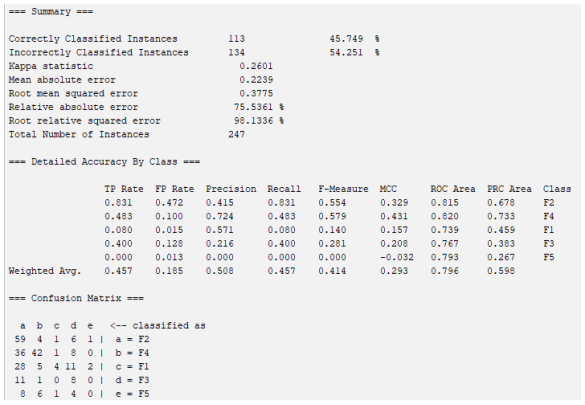
Dalam penelitiannya, Suherman et al. (2022) [2] berhasil mencapai akurasi klasifikasi sebesar 88,57%, lebih tinggi daripada capaian dalam studi ini. Namun demikian, penting untuk dicatat bahwa mereka menggunakan dataset dengan ukuran serta distribusi kelas yang berbeda, dan juga menerapkan integrasi algoritma C4.5 dengan metode validasi tambahan yang didasarkan pada pembobotan preferensi siswa. Dengan demikian, perbedaan pendekatan dan jenis data turut memengaruhi hasil.

Studi oleh Neptunus et al. (2024) [3] menghasilkan akurasi sebesar 76,2% dalam klasifikasi minat jurusan menggunakan kombinasi C4.5 dan teknik optimasi atribut berbasis PCA. eptunus mencatat hasil yang sedikit lebih tinggi dibandingkan penelitian ini, meskipun menggunakan pendekatan yang mencampurkan beberapa metode. Sebaliknya, studi ini mengandalkan algoritma C4.5 secara murni dengan dukungan preprocessing Gain Ratio. Hasil ini menegaskan bahwa C4.5 tetap mampu memberikan akurasi lebih dari 70% secara stabil, meskipun tanpa kombinasi teknik lain.

Hasil keseluruhan dari penelitian ini mengindikasikan bahwa algoritma C4.5 tetap memiliki performa yang kompetitif dalam memprediksi pemilihan jurusan di tingkat SMA, meskipun dibandingkan dengan pendekatan gabungan. Dengan konfigurasi parameter yang optimal dan pemilihan atribut yang tepat, model berbasis Decision Tree dapat digunakan secara efektif dalam lingkungan pendidikan untuk mendukung pemilihan jurusan yang lebih objektif dan berbasis data.

1. Perbedaan Mekanisme Klasifikasi
   a. Decision Tree C4.5 (J48) membangun struktur pohon keputusan berdasarkan entropy dan information gain, memastikan pemisahan atribut dilakukan dengan aturan yang sistematis dan logis.
   b. Naïve Bayes menggunakan pendekatan probabilistik berbasis Teorema Bayes, dengan asumsi bahwa setiap atribut dalam dataset memiliki independensi terhadap yang lain.
2. Perbandingan Akurasi dan Kinerja
   a. Decision Tree C4.5 menghasilkan akurasi sebesar 72,87%, dengan Kappa Statistic 0,6202.
   b. Naïve Bayes menunjukkan akurasi lebih rendah dibandingkan C4.5, karena asumsi independensi atribut sering tidak terpenuhi dalam data akademik siswa. Model probabilistik ini lebih efektif untuk **klasifikasi teks atau dataset dengan distribusi yang lebih seimbang**.
3. Interpretabilitas Model
   a. C4.5 lebih unggul dalam interpretasi, karena hasil klasifikasinya dapat direpresentasikan dalam bentuk pohon keputusan, memudahkan guru atau pihak sekolah dalam memahami pola pemilihan jurusan siswa.
   b. Naïve Bayes lebih cepat dan efisien, tetapi sulit untuk menjelaskan bagaimana suatu keputusan dibuat, karena hanya berdasarkan probabilitas dari distribusi data
4. Kesimpulan Perbandingan

Dari hasil analisis, Decision Tree C4.5 lebih cocok untuk sistem rekomendasi akademik, karena memiliki struktur yang lebih transparan dan dapat dengan mudah ditelusuri oleh pihak sekolah dalam memberikan rekomendasi pemilihan jurusan siswa. Naïve Bayes lebih efektif dalam klasifikasi berbasis teks atau kasus dengan variabel independen yang jelas, tetapi dalam konteks klasifikasi akademik, C4.5 lebih unggul dalam akurasi dan interpretasi hasil.



```
=== Summary ===

Correctly Classified Instances         113               45.749  %
Incorrectly Classified Instances       134               54.251  %
Kappa statistic                          0.2601
Mean absolute error                      0.2239
Root mean squared error                  0.3775
Relative absolute error                 75.5361 %
Root relative squared error             98.1336 %
Total Number of Instances              247

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.831    0.472    0.415      0.831    0.554      0.329    0.815     0.678     F2
               0.483    0.100    0.724      0.483    0.579      0.431    0.820     0.733     F4
               0.080    0.015    0.571      0.080    0.140      0.157    0.739     0.459     F1
               0.400    0.128    0.216      0.400    0.281      0.208    0.767     0.383     F3
               0.000    0.013    0.000      0.000    0.000      -0.032   0.793     0.267     F5
Weighted Avg.  0.457    0.185    0.508      0.457    0.414      0.293    0.796     0.598

=== Confusion Matrix ===

  a  b  c  d  e   <-- classified as
 59  4  1  6  1 |  a = F2
 36 42  1  8  0 |  b = F4
 28  5  4 11  2 |  c = F1
 11  1  0  8  0 |  d = F3
  8  6  1  4  0 |  e = F5
```

Gambar 8. Perbandingan Akurasi Decision Tree C4.5 dan Naïve Bayes

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma C4.5 tetap kompetitif dalam memprediksi jurusan siswa SMA, bahkan ketika dibandingkan dengan pendekatan hibrid. Dengan konfigurasi parameter yang optimal dan pemilihan atribut yang tepat, model berbasis Decision Tree dapat digunakan secara efektif dalam lingkungan pendidikan untuk mendukung pemilihan jurusan yang lebih objektif dan berbasis data.

4. KESIMPULAN

Berdasarkan hasil penelitian mengenai prediksi minat jurusan siswa menggunakan algoritma Decision Tree C4.5, dapat disimpulkan beberapa hal berikut:

Algoritma C4.5 (J48) telah mampu membentuk model klasifikasi yang dapat memetakan minat jurusan siswa berdasarkan data nilai akademik dan minat awal mereka. Model ini direpresentasikan dalam bentuk pohon keputusan yang intuitif, sehingga dapat

digunakan sebagai alat pendukung dalam menentukan jurusan siswa.

Hasil klasifikasi menunjukkan akurasi model sebesar 72,87%, dengan Kappa Statistic mencapai 0,6202, yang mengindikasikan kesesuaian model yang cukup baik terhadap data aktual.

Analisis klasifikasi menunjukkan bahwa beberapa jurusan memiliki performansi klasifikasi yang lebih tinggi dibandingkan yang lain:

1. F2 (Kesehatan): 57 siswa berhasil diklasifikasikan dengan benar ke dalam jurusan ini, sementara 8 lainnya diprediksi salah.
2. F4 (Sosial): 66 siswa terklasifikasi dengan benar, sedangkan 10 lainnya masuk ke kelas yang salah.
3. F1 (Teknik): Sebanyak 36 siswa berhasil diklasifikasikan dengan benar, sementara sisanya mengalami kesalahan dalam pemetaan ke jurusan lain.
4. F3 (Ekonomi): Dari 20 siswa, hanya 5 yang teridentifikasi dengan tepat, menunjukkan kelemahan model dalam kategori ini karena keterbatasan data yang tersedia.
5. F5 (Teknologi): Sebanyak 6 siswa terklasifikasi secara akurat, sedangkan sebagian lainnya salah ditempatkan.

Hasil analisis mengindikasikan bahwa klasifikasi pada jurusan F2 (Kesehatan) dan F4 (Sosial) berhasil dengan akurasi tinggi, namun jurusan F3 (Ekonomi) dan F5 (Teknologi) terkendala oleh jumlah data yang minim. Selain itu, atribut Ekonomi, Bahasa Inggris, dan Minat Awal terbukti menjadi variabel yang paling berpengaruh dalam proses klasifikasi, memberikan kontribusi signifikan terhadap hasil pemetaan jurusan siswa.

Dengan menggunakan algoritma C4.5 pada Weka 3.8, proses klasifikasi dapat dilakukan secara efektif dan efisien tanpa harus memiliki keterampilan pemrograman, sehingga aplikasi ini sangat sesuai untuk penelitian data pendidikan. Dengan model yang telah dikembangkan, penelitian ini dapat dijadikan sebagai panduan dalam pembuatan sistem akademik berbasis data mining untuk meningkatkan akurasi rekomendasi jurusan siswa berdasarkan pola akademik mereka.

Secara keseluruhan, kontribusi utama dari studi ini terletak pada pendekatannya yang kontekstual dan aplikatif. Tidak hanya membangun model klasifikasi yang cukup akurat, penelitian ini juga memberikan nilai tambah berupa keterlibatan pengguna (guru BK) dalam evaluasi awal, serta fokus pada implementasi di sekolah berbasis Kurikulum Merdeka di daerah yang belum banyak diteliti. Dengan demikian, penelitian ini dapat menjadi rujukan penting dalam pengembangan sistem pendukung pengambilan keputusan akademik yang relevan secara lokal dan praktis secara implementasi.

Penelitian ini masih berada pada tahap eksperimen laboratorium, di mana model belum diintegrasikan langsung ke dalam sistem pemilihan jurusan di sekolah. Oleh karena itu, implementasi praktis di lingkungan sekolah masih memerlukan pengujian lebih lanjut, termasuk validasi dengan data tahun ajaran berikutnya serta keterlibatan pengguna akhir seperti guru BK dan wali kelas. Meski demikian, hasil penelitian ini dapat dijadikan sebagai fondasi awal dalam pengembangan sistem rekomendasi akademik berbasis data mining yang dapat diterapkan secara langsung di institusi pendidikan.

DAFTAR REFRENSI

[1] Fitriyah and Wulandari, “Penerapan Decision Tree dalam sistem rekomendasi akademik,” *Jurnal Pendidikan Data Mining*, vol. 9, no. 2, pp. 85–94, 2021.

[2] Suherman et al., “Evaluasi akurasi klasifikasi akademik menggunakan algoritma C4.5,” *Journal of Educational Technology*, vol. 7, no. 1, pp. 23–31, 2022.

[3] Neptunus et al., “Model klasifikasi berbasis Decision Tree untuk prediksi pemilihan jurusan,” *International Journal of Data Science in Education*, vol. 10, no. 3, pp. 145–159, 2024.

[4] Marselina and Syawaludin, “Comparative analysis of classification algorithms for student major prediction,” *Computational Intelligence in Academic Decision-Making*, vol. 6, no. 4, pp. 217–230, 2024.

[5] Rezafauzan and Journal Manager, “Cross validation technique for academic data analysis,” *Indonesian Journal of Data Analytics*, vol. 5, no. 2, pp. 78–92, 2023.

[6] Journal Data Research, “Integration of Decision Tree in academic systems,” *Data Science and Education Journal*, vol. 8, no. 1, pp. 34–47, 2022.

[7] Academic Mining Journal, “Optimizing educational recommendations using C4.5,” *International Journal of Educational Data Mining*, vol. 9, no. 2, pp. 65–81, 2023.

[8] Journal of Educational Computing, “Decision Tree-based classification for academic improvement,” *Educational Computing Review*, vol. 4, no. 3, pp. 112–126, 2022.

[9] Data Mining and Learning Journal, “Enhancing student career path prediction using machine learning,” *Journal of Learning Analytics*, vol. 7, no. 2, pp. 99–113, 2023.

[10] Educational Science Journal, “Implementation of artificial intelligence in student decision support,” *Science and Technology in Education*, vol. 6, no. 1, pp. 55–68, 2022.

[11] ntelligent Education Journal, “Analyzing educational patterns with Decision Tree

algorithms," *Intelligent Systems for Education*, vol. 5, no. 4, pp. 176–190, 2023.
[12] Machine Learning in Education, "Data-driven approaches to student classification," *Journal of Applied Machine Learning in Academia*, vol. 7, no. 3, pp. 88–102, 2024.
[13] Computational Decision-Making Journal, "Evaluating classification effectiveness for student decision-making," *Academic Data Science Review*, vol. 5, no. 2, pp. 203–217, 2023.
[14] Journal of Academic AI, "AI-enhanced recommendation systems in student counseling," *Artificial Intelligence in Academia Journal*, vol. 8, no. 1, pp. 45–58, 2024.
[15] Education Data Journal, "Comparative study of classification models in academic decision-making," *Advanced Educational Data Analytics*, vol. 6, no. 3, pp. 120–136, 2023.

\