

Penerapan Metode *Decision Tree* dalam Klasifikasi Penerimaan Siswa Baru Sebagai Sistem Pendukung Keputusan

Raja Rahmad Hafidzul Manik^{1*}, Rakhmat Kurniawan²

^{1,2}Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ^{1*}Rajarahmad19@gmail.com, ²rakhmat.kr@uinsu.ac.id

(*Email Corresponding Author: Rajarahmad19@gmail.com)

Received: 12 Juni 2026 | Revision: 28 Agustus 2026 | Accepted: 18 September 2026

Abstrak

Penerimaan siswa baru di pesantren melibatkan berbagai kriteria akademik, karakter, keagamaan, dan kedisiplinan, sehingga proses seleksi menjadi menantang ketika evaluasi dilakukan secara manual. Di Pesantren Darul Arafah Medan, data historis penerimaan memuat beragam atribut penilaian yang memerlukan analisis sistematis guna mendukung keputusan penerimaan yang objektif dan konsisten. Penelitian ini bertujuan menerapkan metode *Decision Tree* dengan algoritma C4.5 untuk mengklasifikasikan calon siswa ke dalam kategori diterima atau tidak diterima serta mengevaluasi kinerja model klasifikasi yang dihasilkan. Penelitian ini menggunakan pendekatan kuantitatif dan metode data mining berdasarkan 800 data historis calon siswa. Kumpulan data tersebut mencakup 10 kriteria seleksi, antara lain nilai rapor, nilai tes masuk, prestasi, hasil wawancara, penilaian sikap, kompetensi membaca Al-Qur'an, hafalan, dan kehadiran. Data diproses awal, ditransformasikan menggunakan label encoding, serta dibagi menjadi 80% data latih dan 20% data uji. Algoritma C4.5 menentukan struktur keputusan menggunakan Entropy, *Information Gain*, *Split Information*, dan *Gain Ratio*. Hasil evaluasi menunjukkan tingkat akurasi sebesar 71%, presisi 67%, recall 81%, dan F1-score 74%. Model yang dihasilkan diimplementasikan dalam sistem pendukung keputusan berbasis web untuk memberikan klasifikasi, aturan keputusan, dan rekomendasi penerimaan. Penelitian ini menyimpulkan bahwa algoritma C4.5 secara efektif mendukung pengambilan keputusan penerimaan siswa yang objektif, terukur, dan transparan.

Kata Kunci: Algoritma C4.5, Confusion Matrix, Klasifikasi, Pohon Keputusan, Penerimaan Siswa.

Abstract

New student admissions at Islamic boarding schools involve various academic, character, religious, and disciplinary criteria, making the selection process challenging when evaluations are conducted manually. At Pesantren Darul Arafah Medan, historical admission data contains diverse assessment attributes that require systematic analysis to support objective and consistent admission decisions. This study aims to apply the *Decision Tree* method using the C4.5 algorithm to classify prospective students into accepted or rejected categories, as well as to evaluate the performance of the resulting classification model. This study employs a quantitative approach and data mining techniques based on 800 historical records of prospective students. The dataset comprises 10 selection criteria, including report card grades, entrance test scores, achievements, interview results, attitude assessments, Al-Qur'an reading competence, memorization, and attendance. The data was preprocessed, transformed using label encoding, and divided into 80% training data and 20% testing data. The C4.5 algorithm determines the decision structure using Entropy, *Information Gain*, *Split Information*, and *Gain Ratio*. The evaluation results demonstrated an accuracy rate of 71%, a precision of 67%, a recall of 81%, and an F1-score of 74%. The resulting model was implemented into a web-based decision support system to provide classifications, decision rules, and admission recommendations. This study concludes that the C4.5 algorithm effectively supports objective, measurable, and transparent decision-making for student admissions.

Keywords: C4.5 Algorithm, Confusion Matrix, Classification, Decision Tree, Student Admission.

1. PENDAHULUAN

Pendidikan memegang peranan penting dalam meningkatkan kualitas sumber daya manusia serta mempersiapkan peserta didik untuk memenuhi harapan akademik dan kelembagaan dari institusi pendidikan. Kualitas peserta didik yang diterima oleh suatu institusi pendidikan dipengaruhi oleh proses seleksi yang diterapkan pada saat penerimaan siswa baru [1]. Oleh karena itu, proses penerimaan yang efektif harus mampu menilai calon peserta didik secara objektif berdasarkan kriteria akademik maupun non-akademik yang relevan. Pada institusi pendidikan Islam, proses seleksi juga dapat mempertimbangkan kompetensi tambahan, seperti kemampuan membaca Al-Qur'an yang mencakup aspek tajwid, makhraj, dan fasahah. Kondisi-kondisi tersebut menjadikan penerimaan siswa baru sebagai sebuah proses pengambilan keputusan yang melibatkan berbagai atribut penilaian serta memerlukan pengolahan data yang sistematis [2].

Pesantren Darul Arafah Medan merupakan lembaga pendidikan Islam yang menyelenggarakan proses penerimaan santri baru setiap tahunnya. Proses seleksi ini melibatkan berbagai komponen penilaian, antara lain identitas calon santri, nilai rapor, nilai tes masuk, prestasi, hasil wawancara, tes psikologi sikap, serta kemampuan Tajwid, Makhraj, Fasahah, dan hafalan surat-surat Al-Qur'an. Berbagai aspek tersebut mencerminkan kemampuan akademik, kepribadian, dan keagamaan calon santri [3]. Meningkatnya volume dan keragaman data penilaian menuntut adanya metode yang tepat untuk mengolah informasi tersebut serta mendukung pengambilan keputusan penerimaan [4]. Selain itu, ketersediaan data

historis seleksi membuka peluang untuk memanfaatkan teknik *data mining* guna mengidentifikasi pola-pola yang dapat mendukung keputusan penerimaan di masa mendatang [5], [6].

Dalam praktiknya, pemrosesan dan evaluasi data penerimaan santri baru di Pesantren Darul Arafah Medan masih dilakukan secara manual. Panitia seleksi mengevaluasi hasil tes dan menetapkan keputusan penerimaan berdasarkan informasi penilaian yang tersedia. Pendekatan ini dapat memakan waktu yang cukup lama, menimbulkan tidak konsistennan dalam proses evaluasi, serta menyulitkan identifikasi pola dari data penerimaan sebelumnya [1]. Teknik klasifikasi dalam *data mining* dapat digunakan untuk mengelompokkan data ke dalam kelas-kelas yang telah ditentukan sebelumnya berdasarkan pola yang dipelajari dari data terdahulu [7], [8]. Oleh karena itu, penerapan teknik klasifikasi dapat memberikan pendekatan alternatif untuk mendukung proses penerimaan yang lebih sistematis dan berbasis data [2], [9].

Salah satu metode klasifikasi yang dapat diterapkan pada masalah ini adalah algoritma *Decision Tree* (Pohon Keputusan) [10]. *Decision Tree* merupakan metode pembelajaran terawasi (*supervised learning*) yang membangun struktur berbasis pohon dengan memilih atribut yang memberikan pemisahan data paling relevan ke dalam kelas-kelas target. Pohon yang dihasilkan terdiri dari simpul keputusan, cabang, dan simpul daun yang merepresentasikan hasil klasifikasi. Salah satu keunggulan utama *Decision Tree* adalah kemampuannya untuk diinterpretasikan, karena proses klasifikasi yang dihasilkan dapat direpresentasikan sebagai serangkaian aturan keputusan yang mudah dipahami [4]. Karakteristik ini menjadikan metode tersebut cocok untuk aplikasi pendukung keputusan, di mana alasan di balik suatu rekomendasi perlu dipahami oleh para pengambil keputusan [11].

Penelitian sebelumnya telah menunjukkan potensi *Decision Tree* beserta variannya dalam permasalahan seleksi dan penerimaan di bidang pendidikan. Penelitian Aprilia dkk [9] menerapkan pendekatan *Decision Tree* untuk klasifikasi penerimaan mahasiswa baru dan melaporkan bahwa metode tersebut mampu menghasilkan aturan keputusan yang mudah dipahami guna mendukung proses seleksi mahasiswa. Penelitian lain [2] juga menerapkan teknik penambangan data menggunakan *Decision Tree* untuk klasifikasi penerimaan mahasiswa baru dan menunjukkan bahwa algoritma tersebut dapat mengidentifikasi pola keputusan berdasarkan data penilaian mahasiswa. Selain itu, Pasaribu dkk [12] meneliti algoritma C4.5 yang merupakan varian dari *Decision Tree* untuk penerimaan mahasiswa dan menemukan bahwa pendekatan tersebut dapat membantu institusi pendidikan dalam mengambil keputusan penerimaan secara lebih sistematis. Penelitian-penelitian ini mengindikasikan bahwa metode klasifikasi berbasis pohon memiliki potensi untuk mendukung pengambilan keputusan penerimaan [13].

Meskipun penelitian sebelumnya telah menunjukkan kegunaan metode *Decision Tree* untuk klasifikasi penerimaan siswa, perbedaan dalam atribut penilaian dan kriteria seleksi institusi membuka peluang bagi penelitian lebih lanjut. Penelitian yang ada umumnya menggunakan atribut akademik dan penerimaan konvensional, sedangkan penelitian ini menggabungkan atribut akademik, non-akademik, psikologis, prestasi, dan kompetensi Al-Qur'an termasuk Tajwid, Makhraj, Fasohah, dan hafalan Al-Qur'an dalam konteks pesantren. Kombinasi ini menghadirkan konteks penelitian khas yang belum banyak dieksplorasi dalam studi klasifikasi penerimaan siswa sebelumnya. Oleh karena itu, penelitian ini mengisi celah penelitian tersebut dengan menerapkan dan mengevaluasi metode *Decision Tree* menggunakan data penilaian multidimensi dari Pesantren Darul Arafah Medan. Kebaruan penelitian ini terletak pada penggunaan atribut seleksi spesifik tersebut untuk mengidentifikasi pola klasifikasi dan menghasilkan rekomendasi penerimaan berdasarkan data historis, sekaligus memosisikan model yang dihasilkan sebagai alat pendukung keputusan yang membantu, bukan menggantikan, peran panitia penerimaan.

Berdasarkan permasalahan yang teridentifikasi dan kesenjangan penelitian, studi ini bertujuan untuk menerapkan metode *Decision Tree* guna mengklasifikasikan calon siswa, menghasilkan rekomendasi penerimaan, serta mengevaluasi kinerja klasifikasi model tersebut. Hasil yang diharapkan adalah terciptanya pendekatan analisis yang objektif dan konsisten yang dapat mendukung pemanfaatan data historis penerimaan dalam proses seleksi siswa.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode Data Mining untuk membangun model klasifikasi penerimaan siswa baru. Metode yang digunakan adalah *Decision Tree* dengan algoritma C4.5. Penelitian bertujuan menghasilkan model klasifikasi yang dapat membantu proses pengambilan keputusan penerimaan calon siswa baru berdasarkan karakteristik dan kriteria yang telah ditentukan oleh Pesantren Darul Arafah Medan.



Gambar 1. Alur tahapan penelitian

Tahapan penelitian meliputi perencanaan, pengumpulan data, analisis data, penerapan model *Decision Tree* menggunakan algoritma C4.5, dan pengujian model. Secara umum, alur penelitian dirancang dari data calon siswa sebagai masukan hingga menghasilkan keputusan berupa kategori diterima atau tidak diterima.

a. Perencanaan Penelitian

Proses penelitian ini diawali dengan perencanaan yaitu menentukan topik yang akan dibahas. Topik penelitian ini yaitu penerapan metode *Decision Tree* dalam klasifikasi penerimaan siswa baru sebagai sistem pendukung keputusan (studi kasus Pesantren Darul Arafah Medan).

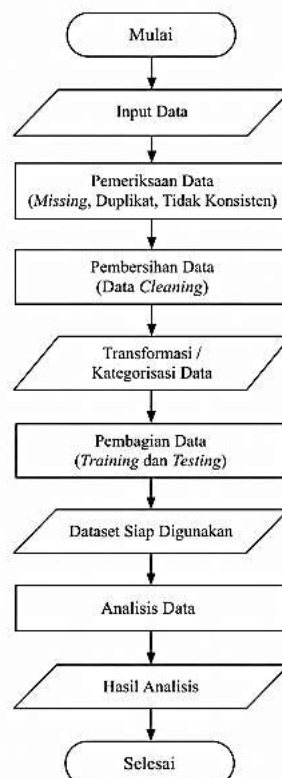
b. Pengumpulan Data

Penelitian ini dilakukan dengan cara observasi langsung ke sekolah Pesantren Darul Arafah Medan, untuk mengetahui data penerimaan siswa baru dan syarat apa yang harus dilalui ketika proses penerimaan siswa baru. Teknik ini dilakukan dengan mengamati objek yang menjadi fokus penelitian, baik itu berupa manusia, lingkungan, atau sistem yang digunakan dalam pengambilan keputusan. Selanjutnya dilakukan wawancara langsung dengan Kepala Sekolah dimana data diperoleh dari catatan akademik sekolah atau diberikan langsung oleh pihak sekolah.

c. Analisis Data

Data yang diperoleh dari sekolah terlebih dahulu dianalisis untuk menentukan atribut yang relevan dengan proses klasifikasi. Atribut yang digunakan meliputi nilai rapor, nilai tes masuk, prestasi, hasil wawancara, dan status penerimaan siswa. Klasifikasi merupakan proses pengelompokan data ke dalam kelas tertentu berdasarkan karakteristik yang dimiliki. Dalam penelitian ini, klasifikasi digunakan untuk menentukan status calon siswa ke dalam dua kelas, yaitu diterima dan tidak diterima, sesuai dengan konsep *binary classification* [8].

Selanjutnya dilakukan data *preprocessing* yang meliputi pemeriksaan, pembersihan, dan transformasi data. Pemeriksaan dilakukan untuk mengetahui adanya data kosong, data ganda, dan data yang tidak konsisten [14]. Data yang bermasalah kemudian diperbaiki atau dihapus sesuai kebutuhan. Selanjutnya, data numerik dapat ditransformasikan atau dikategorikan sesuai dengan kebutuhan penerapan algoritma C4.5. Setelah data siap digunakan, dataset dibagi menjadi data *training* dan *testing*. Berikut merupakan rancangan yang dilakukan peneliti berupa *flowchart* di bawah ini:



Gambar 2. Flowchart Metode *Decision Tree* C4.5

Tahapan penelitian diawali dengan memasukkan dan memeriksa data calon siswa untuk memastikan tidak terdapat data kosong, duplikat, atau tidak konsisten. Data kemudian dibersihkan dan ditransformasikan sesuai kebutuhan, selanjutnya dibagi menjadi data *training* dan *testing* dengan teknik *Stratified Train-Test Split* untuk menjaga proporsi seimbang pada setiap kelas target [15]. Data *training* digunakan untuk membangun model, sedangkan data *testing* digunakan untuk menguji hasil klasifikasi. Selanjutnya, algoritma C4.5 diterapkan dengan

menghitung *Entropy*, *Information Gain*, *Split Information*, dan *Gain Ratio* untuk menentukan atribut terbaik sebagai simpul pohon keputusan. Model yang terbentuk menghasilkan *decision rules* untuk mengklasifikasikan calon siswa menjadi diterima atau tidak diterima, kemudian kinerjanya dievaluasi menggunakan data testing.

d. Penerapan *Decision Tree* C4.5

Metode *Decision Tree* digunakan untuk membentuk model klasifikasi dalam bentuk struktur pohon keputusan yang terdiri dari *root node*, *branch*, *internal node*, dan *leaf node*. Pada penelitian ini, *leaf node* menghasilkan keputusan akhir berupa diterima atau tidak diterima.

Algoritma yang digunakan untuk membentuk pohon keputusan adalah C4.5. Algoritma C4.5 menentukan atribut terbaik berdasarkan nilai *Gain Ratio*. Proses pembentukan pohon dilakukan secara bertahap dengan menghitung *Entropy*, *Information Gain*, *Split Information*, dan *Gain Ratio* pada setiap atribut [16]. Atribut yang memiliki nilai *Gain Ratio* tertinggi dipilih sebagai simpul (*node*) dan proses pembagian data dilakukan secara berulang sampai diperoleh keputusan akhir [17].

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

Keterangan :

(S) = himpunan data,

(p_i) = proporsi data pada kelas ke-i, dan

(n) = jumlah kelas.

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

Keterangan :

(S) = himpunan data,

(A) = atribut,

(S_v) = subset data berdasarkan nilai atribut (v), dan

(|S|) = jumlah keseluruhan data.

$$SplitInfo(S, A) = -\sum_{v \in Values(A)} \frac{|S_v|}{|S|} \log_2 \left(\frac{|S_v|}{|S|} \right) \quad (3)$$

Nilai *Split Information* digunakan bersama *Information Gain* dalam menentukan nilai *Gain Ratio*.

Gain Ratio merupakan ukuran yang digunakan oleh algoritma C4.5 untuk menentukan atribut terbaik sebagai simpul pada *Decision Tree*. Rumus yang digunakan adalah:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \quad (4)$$

Atribut dengan nilai *Gain Ratio* tertinggi dipilih sebagai *root node* atau simpul pada proses pembentukan pohon keputusan. Perhitungan tersebut dilakukan secara berulang pada setiap subset data hingga terbentuk pohon keputusan yang menghasilkan kelas akhir.

e. Pengujian

Pengujian sistem dilakukan untuk mengetahui tingkat akurasi model *Decision Tree* dalam mengklasifikasikan data calon siswa baru. Pengujian dilakukan dengan membagi data menjadi data *training* dan data *testing*, yaitu 80% data *training* dan 20% data *testing*. Data *training* digunakan untuk membangun model, sedangkan data *testing* digunakan untuk menguji kemampuan model dalam melakukan prediksi. Hasil pengujian dianalisis menggunakan metode *Confusion Matrix* untuk mengetahui jumlah prediksi yang benar dan salah. Selanjutnya dilakukan perhitungan nilai akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur kinerja model [18]. Semakin tinggi nilai akurasi yang diperoleh, maka semakin baik model dalam melakukan klasifikasi [7].

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk membangun sebuah Sistem Pendukung Keputusan (SPK) yang objektif, terukur, serta mampu mengakomodasi pilar-pilar penilaian holistik khas lembaga pendidikan pesantren, meliputi aspek akademik, karakter, kedisiplinan, hingga kemampuan baca dan hafalan Al-Qur'an.

3.1 Pembahasan

Analisis data merupakan tahapan krusial dalam ekstraksi pengetahuan (*knowledge discovery*) dari dataset historis pendaftaran calon santri baru. Proses ini bertujuan untuk memahami karakteristik variabel, memastikan kualitas serta keutuhan data, serta mengidentifikasi pola distribusi data sebelum ditransformasikan ke dalam bentuk model pohon keputusan. Dataset yang digunakan dalam penelitian ini terdiri dari 800 data historis calon santri baru. Setiap entitas

pendaftar diwakili oleh 10 fitur seleksi (fitur prediktor), 1 identifikasi siswa (ID_Siswa), serta 1 atribut label target (Status).

Penentuan 10 fitur seleksi disesuaikan dengan kriteria evaluasi penerimaan santri di Pesantren Darul Arafah Medan yang mencakup empat pilar penilaian, yaitu akademik, karakter dan non-akademik, keagamaan dan Al-Qur'an, serta kedisiplinan. Pilar akademik meliputi Nilai_Rapor dan Nilai_Tes_Masuk, pilar karakter dan non-akademik meliputi Hasil_Wawancara, Psikotes_Sikap, dan Prestasi, pilar keagamaan dan Al-Qur'an meliputi Tajwid, Makhraj, Fasohah, dan Hafalan_Juz, sedangkan pilar kedisiplinan direpresentasikan oleh Status_Kehadiran. Keseluruhan fitur tersebut digunakan sebagai variabel prediktor untuk menentukan kelas target Status, yaitu Diterima atau Tidak Diterima, dengan rincian struktur variabel dataset disajikan pada tabel 1 berikut.

Tabel 1. Deskripsi Atribut Dataset Penerimaan Santri Baru

Nama Atribut	Peran dalam Model	Pilar Evaluasi	Deskripsi Variabel	Skala / Domain Nilai
ID_Siswa	Identifier	Identitas	Kode unik pengenalan calon santri	String (contoh: CSB-001)
Nilai_Rapor	Fitur (Prediktor)	Akademik	Rata-rata nilai rapor sekolah asal	Numerik Kontinu (50.0 - 98.0)
Nilai_Tes_Masuk	Fitur (Prediktor)	Akademik	Nilai ujian tulis potensi akademik	Numerik Kontinu (40.0 - 100.0)
Prestasi	Fitur (Prediktor)	Karakter	Kepemilikan sertifikat prestasi	Kategorikal (Ada, Tidak Ada)
Hasil_Wawancara	Fitur (Prediktor)	Karakter	Evaluasi wawancara tatap muka	Ordinal (Sangat Baik, Baik, Cukup, Kurang)
Psikotes_Sikap	Fitur (Prediktor)	Karakter	Penilaian kepribadian & psikologi	Ordinal (Sangat Baik, Baik, Cukup, Kurang)
Tajwid	Fitur (Prediktor)	Keagamaan	Ketepatan hukum bacaan Al-Qur'an	Ordinal (Sangat Baik, Baik, Cukup, Kurang)
Makhraj	Fitur (Prediktor)	Keagamaan	Ketepatan pelafalan huruf Al-Qur'an	Ordinal (Sangat Baik, Baik, Cukup, Kurang)
Fasohah	Fitur (Prediktor)	Keagamaan	Kelancaran dan fasih membaca Al-Qur'an	Ordinal (Sangat Baik, Baik, Cukup, Kurang)
Hafalan_Juz	Fitur (Prediktor)	Keagamaan	Jumlah hafalan Al-Qur'an (Tahfidz)	Ordinal (> 3 Juz, 1 - 2 Juz, 0 Juz)
Status_Kehadiran	Fitur (Prediktor)	Kedisiplinan	Status partisipasi jadwal ujian	Kategorikal (Hadir Lengkap, Susulan)
Status	Target Class	Keputusan SPK	Keputusan akhir penerimaan santri	Binary (Diterima, Tidak Diterima)

Kemudian adalah analisis pembersihan data dimana kualitas model keputusan yang dihasilkan oleh algoritma C4.5 sangat bergantung pada kebersihan (*cleanness*) dan keutuhan data input. Adanya nilai yang hilang (*missing values*) atau data ganda (*duplicate rows*) dapat membiaskan perhitungan nilai *Entropy* dan *Information Gain*, yang berpotensi menghasilkan struktur pohon keputusan yang tidak akurat. Hasil pemeriksaan menunjukkan bahwa tidak ditemukan satu pun kolom atau baris yang mengalami kekosongan data.

Tabel 2. Hasil Pengujian Missing Value per Atribut

Nama Atribut	Jumlah Data Terisi	Jumlah Missing Value	Persentase Keutuhan Data
ID_Siswa	800	0	100%
Nilai_Rapor	800	0	100%
Nilai_Tes_Masuk	800	0	100%
Prestasi	800	0	100%
Hasil_Wawancara	800	0	100%
Psikotes_Sikap	800	0	100%
Tajwid	800	0	100%
Makhraj	800	0	100%
Fasohah	800	0	100%
Hafalan_Juz	800	0	100%
Status_Kehadiran	800	0	100%
Status	800	0	100%

Berdasarkan Tabel 2, tingkat keutuhan dataset mencapai 100% (seluruh 800 entri lengkap), sehingga tidak diperlukan prosedur imputation (pengisian data hilang) maupun penanganan *missing value* khusus. Langkah selanjutnya adalah pemeriksaan duplikasi data, dimana hasil pengujian algoritma verifikasi menunjukkan jumlah 0 baris duplikat (0%). Dengan demikian, dataset dipastikan bersih dari redundansi data dan siap digunakan untuk proses pemodelan selanjutnya.

Kemudian Algoritma *Decision Tree* C4.5 menangani dua jenis tipe data yang berbeda, yaitu data numerik (kontinu) dan data kategorikal (nominal/ordinal). Untuk data numerik (kontinu), dataset terdiri atas dua fitur numerik (kontinu), yaitu Nilai_Rapor dan Nilai_Tes_Masuk, serta delapan fitur kategorikal yang digunakan sebagai prediktor. Statistik deskriptif fitur numerik ditunjukkan pada tabel 3 berikut.

Tabel 3. Ringkasan Statistik Deskriptif Fitur Numerik (N=800)

Parameter Statistik	Nilai_Rapor	Nilai_Tes_Masuk
Rata-rata (Mean)	77.88	75.98
Simpangan Baku (Std Dev)	8.70	11.87
Nilai Minimum (Min)	50.00	40.00
Kuartil Bawah (Q1 / 25%)	71.70	68.00
Median (Q2 / 50%)	78.10	75.90
Kuartil Atas (Q3 / 75%)	83.70	83.80
Nilai Maksimum (Max)	98.00	100.00

Berdasarkan Tabel 3, dapat dilihat bahwa rata-rata Nilai_Rapor pendaftar adalah 77.88 dengan standar deviasi 8.70, sementara rata-rata Nilai_Tes_Masuk berada pada 75.98 dengan standar deviasi 11.87. Rentang yang cukup lebar pada nilai tes masuk (40.00 hingga 100.00) memberikan daya pembeda (*discriminative power*) yang sangat kuat bagi model untuk memisahkan calon santri berprestasi tinggi dari yang kurang memenuhi standar. Kemudian algoritma menggunakan fitur kategorikal (nominal dan ordinal) dimana distribusi frekuensi dan persentase untuk seluruh variabel kategorikal ini dirangkum dalam Tabel 4.

Tabel 4. Distribusi Frekuensi dan Persentase Atribut Kategorikal

Nama Atribut	Kategori / Nilai	Jumlah Santri (Frekuensi)	Persentase (%)
Prestasi	Tidak Ada	520	65.00%
	Ada	280	35.00%
	Baik	410	51.25%
Hasil_Wawancara	Cukup	172	21.50%
	Sangat Baik	151	18.88%
	Kurang	67	8.38%
	Baik	380	47.50%
Psikotes_Sikap	Sangat Baik	208	26.00%
	Cukup	138	17.25%
	Kurang	74	9.25%
-	-	-	-
	-	-	-
	-	-	-
	-	-	-
Hafalan_Juz	1 - 2 Juz	426	53.25%
	0 Juz	209	26.13%
	> 3 Juz	165	20.63%
Status_Kehadiran	Hadir Lengkap	691	86.38%
	Susulan	109	13.63%
Status (Target)	Diterima	404	50.50%
	Tidak Diterima	396	49.50%

Dari Tabel 4, beberapa poin penting terkait karakteristik calon santri Pesantren Darul Arafah Medan yang dapat disimpulkan antara lain adalah penyebaran kelas yang sangat seimbang. Dari total 800 calon santri, sebanyak 404 santri (50.50%) dinyatakan Diterima dan 396 santri (49.50%) dinyatakan Tidak Diterima. Kondisi dataset yang sangat seimbang (*balanced dataset*) ini ideal dalam pengujian algoritma *machine learning*, karena mencegah model berprasangka (bias) terhadap salah satu kelas tertentu [13].

Selanjutnya Sebagian besar calon santri telah memiliki dasar bacaan Al-Qur'an yang cukup baik, di mana kategori 'Baik' dan 'Sangat Baik' mendominasi fitur Tajwid (69.63%), Makhraj (70.26%), dan Fasohah (66.13%). Dari sisi hafalan, 53.25% santri memiliki hafalan 1 - 2 Juz, dan 20.63% telah menghafal > 3 Juz serta mayoritas pendaftar yaitu 86.38% (691 santri) menghadiri jadwal seleksi secara lengkap (Hadir Lengkap), sementara 13.63% (109 santri) mengikuti jalur ujian susulan.

Setelah melalui proses tersebut, data teks kategorikal ditransformasikan menjadi bentuk diskrit numerik melalui teknik *Label Encoding*, serta dilakukan pembagian dataset (*data splitting*) untuk kebutuhan pelatihan dan pengujian model [19]. Sebagai bahan ilustrasi dan acuan analisis eksperimental pada tahapan selanjutnya, diambil 10 sampel alternatif calon santri secara acak (*stratified random sampling*) dari keseluruhan dataset historis penerimaan santri baru.

Sampel ini mencakup kualifikasi siswa dari kode CSB-001 hingga CSB-010. Pemilihan 10 data sampel ini dirancang khusus agar mampu memetakan keberagaman (heterogenitas) pola data yang riil di lapangan seperti yang ditunjukkan pada tabel 5.

Tabel 5. Sampel Data

ID Siswa	Nilai Rapor	Nilai Tes Masuk	Prestasi	Hasil Wawancara	Psikotes Sikap	Tajwid	Makhraj	Fasohah	Hafalan Juz	Status Kehadiran
CSB-001	82.5	86.3	Tidak Ada	Kurang	Sangat Baik	Kurang	Baik	Baik	0 Juz	Hadir Lengkap
CSB-002	76.8	68.8	Tidak Ada	Baik	Sangat Baik	Sangat Baik	Sangat Baik	Baik	> 3 Juz	Hadir Lengkap
CSB-003	83.8	76.2	Ada	Baik	Baik	Baik	Kurang	Baik	1 - 2 Juz	Hadir Lengkap
-	-	-	-	-	-	-	-	-	-	-
CSB-010	82.9	64.3	Tidak Ada	Kurang	Sangat Baik	Kurang	Sangat Baik	Baik	> 3 Juz	Hadir Lengkap

Setelah melewati proses perhitungan manual, skema *Label Encoding* bekerja dengan cara memetakan setiap string/kategori unik dari suatu variabel secara konsisten ke dalam bentuk matriks bilangan bulat (integer) berurutan (mulai dari 0,1,2, dan seterusnya). Rincian pemetaan (*mapping dictionary*) untuk setiap variabel kategorikal beserta kode numerik hasil transformasinya disajikan secara lengkap pada Tabel 6.

Tabel 6. Aturan Pemetaan Label Encoding per Atribut Kategorikal

Nama Atribut	Nilai Kategori Asli (Text)	Nilai Ter-encode (Integer)
Prestasi	Ada	0
	Tidak Ada	1
	Baik	0
Hasil_Wawancara	Cukup	1
	Kurang	2
	Sangat Baik	3
	Baik	0
Psikotes_Sikap	Cukup	1
	Kurang	2
	Sangat Baik	3
-	-	-
	-	-
Status (Target)	Diterima	0
	Tidak Diterima	1

Dengan mengacu pada aturan pemetaan pada Tabel 6, seluruh data mentah pada Tabel 5 mengalami proses encoding dimana hasil transformasi matriks numerik untuk 10 sampel alternatif disajikan pada Tabel 7.

Tabel 7. Hasil Transformasi Label Encoding

ID Siswa	Nilai Rapor	Nilai Tes	Prestasi	Wawancara	Psikotes	Tajwid	Makhraj	Fasohah	Hafalan	Kehadiran	Status
CSB-001	82.5	86.3	1	2	3	2	0	0	0	0	1
CSB-002	76.8	68.8	1	0	3	3	3	0	2	0	0
CSB-003	83.8	76.2	0	0	0	0	2	0	1	0	0
-	-	-	-	-	-	-	-	-	-	-	-
CSB-010	82.9	64.3	1	2	3	2	3	0	2	0	1

Untuk menguji keandalan dan daya generalisasi model *Decision Tree*, dataset keseluruhan berjumlah 800 data santri dibagi menjadi 2 (dua) bagian independen dengan rasio pembagian 80% : 20%. Pembagian data dilakukan menggunakan metode *Stratified Train-Test Split* dengan parameter stratify=y dan random_state=42. Penggunaan metode *stratified* menjamin bahwa proporsi distribusi kelas target (Diterima dan Tidak Diterima) pada data latih maupun data uji tetap

terjaga seimbang dan sama persis dengan proporsi awal dataset. Rincian pembagian data ditunjukkan pada Tabel 8 berikut.

Tabel 8. Rincian Pembagian Dataset Latih dan Dataset Uji

Kelompok Data	Rasio Pembagian	Kategori Diterima (0)	Kategori Tidak Diterima (1)
Data Latih (<i>Training</i>)	80%	323 santri (50.47%)	317 santri (49.53%)
Data Uji (<i>Testing</i>)	20%	81 santri (50.62%)	79 santri (49.38%)
Total Dataset	100%	404 santri (50.50%)	396 santri (49.50%)

Untuk perhitungan manual langkah awal dalam C4.5 adalah menghitung nilai ketidakmurnian total (Entropy) dari seluruh sampel sistem S ($|S|=10$) sebelum dilakukan pemisahan oleh fitur apa pun. Rincian komponen variabel perhitungan Entropy Total S dijabarkan pada Tabel 9.

Tabel 9. Rincian Komponen Perhitungan Entropy Total Sistem (S)

Kelas Target (i)	Jumlah Santri (Ni)	Probabilitas ($p_i = S_i /N$)	Nilai $\log_2(p_i)$	Komponen $-p_i \log_2(p_i)$	Entropy Total Entropy(S)
Diterima	6	$6/10 = 0.6000$	-0.73697	0.44218	
Tidak Diterima	4	$4/10 = 0.4000$	-1.32193	0.52877	
Total Sistem (S)	10	1.0000	-	$0.44218 + 0.52877$	0.9710

Selanjutnya, perhitungan nilai Entropy bagian ("Entropy" (S_v)) untuk setiap kategori/kelompok pada fitur disajikan pada Tabel 10.

Tabel 10. Rincian Perhitungan Entropy Sub-Kategori Atribut Kategorikal

Nama Atribut	Nilai Kategori (v)	Daftar ID Siswa Anggota	Total (Nv)	Diterima	Tidak Diterima	$p(\text{Diterima})$	$P(\text{Tdk Diterima})$	Entropy Sub-kategori Entropy(S_v)
Hasil_Wawancara	Kurang	CSB-001, CSB-010	2	0	2	0.0000	1.0000	0.0000
	Cukup	CSB-006, CSB-009	2	2	0	1.0000	0.0000	0.0000
	Baik	CSB-002, CSB-003, CSB-004, CSB-005, CSB-007, CSB-008	6	4	2	0.6667	0.3333	0.9183

Khusus atribut numerik kontinu seperti Nilai_Rapor dan Nilai_Tes_Masuk, C4.5 mengurutkan nilai dan menguji kandidat *threshold split* berdasarkan perubahan label target. Setiap kandidat dievaluasi menggunakan Entropy dan *Information Gain* untuk menentukan titik pemisah optimal antara kelas Diterima dan Tidak Diterima. Hasil evaluasi *threshold split* disajikan pada Tabel 11.

Tabel 11. Evaluasi Titik Pemisah (*Threshold Split*) Atribut Numerik Kontinu

Fitur Numerik	Urutan Nilai & Status Target	Candidate Threshold (T)	Kiri ($\leq T$) (N,Dit,TDit)	Entropy Kiri	Kanan ($> T$) (N,Dit,TDit)	Entropy Kanan	Weighted Entropy	Information Gain	Status Kualifikasi
Nilai_Rapor	73.8 (D),	79.65	(4, 4, 0)	0.0000	(6, 2, 4)	0.9183	0.5510	0.4200	Terbaik (Optimal)
	75.9 (D),	83.35	(6, 4, 2)	0.9183	(4, 2, 2)	1.0000	0.9510	0.0200	Non-Optimal
	75.9 (D),	84.35	(7, 5, 2)	0.8631	(3, 1, 2)	0.9183	0.8797	0.0913	Non-Optimal
	76.8 (D),								
	82.5 (T),	91.95	(9, 5, 4)	0.9911	(1, 1, 0)	0.0000	0.8920	0.0790	Non-Optimal
	82.9 (T),								

Fitur Numerik	Urutan Nilai & Status Target	Candidate Threshold (T)	Kiri ($\leq T$) (N,Dit,TDit)	Entropi Kiri	Kanan ($>T$) (N,Dit,TDit)	Entropi Kanan	Weighted Entropy	Information Gain	Status Kualifikasi
	83.8 (D), 84.9 (T), 91.7 (T), 92.2 (D)								

Setelah *Entropy* untuk setiap sub-kategori/kelompok dihitung, langkah berikutnya adalah menghitung *Entropy* Terbobot (*Weighted Entropy*) dan nilai *Information Gain* untuk seluruh 10 atribut pada Node 1 (*Root Node*). Peringkat seluruh 10 atribut berdasarkan perolehan nilai *Information Gain* disajikan pada Tabel 12.

Tabel 12. Matriks Peringkat *Information Gain* Seluruh Atribut pada Node 1 (Akar Utama)

Peringkat	Nama Atribut / Fitur	Tipe Data / Kriteria Split	Entropy Total Sistem Entropy(S)	Entropy Terbobot $\sum \ S\ \ S_v\ \text{Entropy}(S_v)$	Information Gain(S,A)	Keputusan Kualifikasi
1	Hasil_Wawancara	Categorical (Kurang, Cukup, Baik)	0.9710	0.5510	0.4215	Terpilih sebagai Root Node
2	Nilai_Rapor	Numerical (≤ 79.65)	0.9710	0.5510	0.2640	Alternatif Kandidat Utama
3	Tajwid	Categorical (4 Kategori)	0.9710	0.6000	0.2725	Kandidat Simpul Turunan
-	-	-	-	-	-	-
10	Status_Kehadiran	Categorical (Hadir Lengkap, Susulan)	0.9710	0.8920	0.0790	Kandidat Simpul Turunan

Berdasarkan Tabel 12, *Information Gain* tertinggi diperoleh atribut Hasil_Wawancara sebesar 0,4215. Oleh karena itu, atribut tersebut ditetapkan sebagai Root Node yang membagi data ke dalam tiga cabang utama. Struktur cabang dan subset data disajikan pada Tabel 13.

Tabel 13. Pemetaan Status Cabang Utama Hasil Pembagian Root Node (Hasil_Wawancara)

Nilai Cabang Hasil_Wawancara	Anggota ID Siswa Terkait	Jumlah Data (N)	Distribusi Kelas Target
Hasil_Wawancara = Kurang	CSB-001, CSB-010	2	0 Diterima, 2 Tidak Diterima
Hasil_Wawancara = Cukup	CSB-006, CSB-009	2	2 Diterima, 0 Tidak Diterima
Hasil_Wawancara = Baik	CSB-002, CSB-003, CSB-004, CSB-005, CSB-007, CSB-008	6	4 Diterima, 2 Tidak Diterima

Sebagaimana hasil pada Langkah sebelumnya, cabang Hasil_Wawancara = Baik memiliki 6 sampel data (CSB-002, CSB-003, CSB-004, CSB-005, CSB-007, CSB-008) dengan komposisi 4 santri Diterima dan 2 santri Tidak Diterima. Nilai Entropy awal pada Node 1.1 ini adalah :

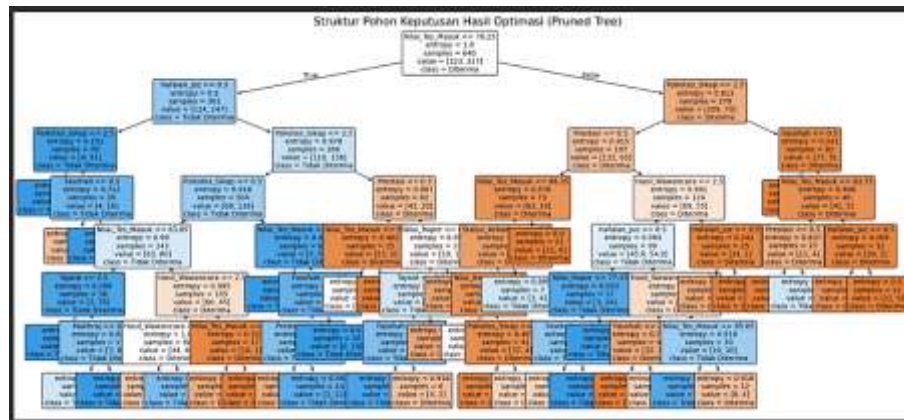
$$\text{Entropy}(S_{\text{Node 1.1}}) = -\left(\frac{4}{6}\right) \log_2 \left(\frac{4}{6}\right) - \left(\frac{2}{6}\right) \log_2 \left(\frac{2}{6}\right) = 0.9183$$

Karena nilai Entropy > 0 , dilakukan perhitungan ulang *Entropy* sub-kategori dan *Information Gain* untuk 9 atribut tersisa pada sub-dataset Node 1.1. Berdasarkan hasil perhitungan dengan 10 data, atribut Tajwid memiliki nilai *Information Gain* tertinggi yaitu 0.5850, sehingga Tajwid ditetapkan sebagai Simpul Pemisah pada Node 1.1. Seluruh hasil pemisahan simpul dari *Root Node* hingga *Leaf Node* dirangkum dalam bentuk hierarki pohon keputusan seperti pada Tabel 14.

Tabel 14. Pemetaan Ringkasan Hierarki Simpul Pohon Keputusan C4.5

Kode Simpul	Tingkat Hierarki	Atribut Evaluasi (Split Criteria)	Nilai Cabang / Syarat	Jumlah Sampel (N)	Distribusi Kelas (Dit / TDiT)	Entropy	Tipe Simpul	Status Target Keputusan
Node 1	Level 0	Hasil_Wawancara	Root Node	10	6 / 4	0.9710	Root Node	-
Leaf 1.1	Level 1	-	Hasil_Wawancara = Kurang	2	0 / 2	0.0000	Leaf Node	Tidak Diterima
Leaf 1.2	Level 1	-	Hasil_Wawancara = Cukup	2	2 / 0	0.0000	Leaf Node	Diterima
Node 1.1	Level 1	Tajwid	Hasil_Wawancara = Baik	6	4 / 2	0.9183	Internal Node	-
Leaf 2.1	Level 2	-	Tajwid = Sangat Baik	2	2 / 0	0.0000	Leaf Node	Diterima
Leaf 2.2	Level 2	-	Tajwid = Cukup	1	0 / 1	0.0000	Leaf Node	Tidak Diterima
Leaf 2.3	Level 2	-	Tajwid = Kurang	1	1 / 0	0.0000	Leaf Node	Diterima
Node 1.1.1	Level 2	Prestasi	Tajwid = Baik	2	1 / 1	1.0000	Internal Node	-
Leaf 3.1	Level 3	-	Prestasi = Ada	1	1 / 0	0.0000	Leaf Node	Diterima
Leaf 3.2	Level 3	-	Prestasi = Tidak Ada	1	0 / 1	0.0000	Leaf Node	Tidak Diterima

Berdasarkan penelusuran dari *Root Node* hingga ke seluruh *Leaf Node* pada Tabel 11, dihasilkan 7 aturan keputusan (rules) logis yang merepresentasikan pengetahuan (*knowledge representation*) dari model C4.5 dimana berhasil mengklasifikasikan seluruh sampel data (10 dari 10 santri) secara tepat tanpa ada kesalahan klasifikasi (*misclassification* = 0), yang membuktikan validitas konsistensi perhitungan manual Algoritma C4.5 pada penelitian ini.



Gambar 3. Potongan Kode dan Visualisasi Grafik Pohon Keputusan Terpankas (*Pruned Decision Tree Plot*)

Setelah melalui proses algoritma *Decision Tree* selanjutnya adalah tahap pengujian *confusion matrix* dimana menggunakan tools seperti *google colab* seperti berikut.

--- LAPORAN EVALUASI PERFORMA AI ---				
	precision	recall	f1-score	support
Diterima	0.67	0.81	0.74	81
Tidak Diterima	0.76	0.59	0.67	79
accuracy			0.71	160
macro avg	0.72	0.70	0.70	160
weighted avg	0.72	0.71	0.70	160

Gambar 4. Visualisasi *Classification Report Model C4.5*

Berdasarkan hasil evaluasi, model memperoleh akurasi 71%, precision 67%, recall 81%, dan F1-score 74% pada kelas Diterima. Hasil tersebut menunjukkan bahwa model memiliki kemampuan klasifikasi yang tinggi dan performa yang relatif seimbang dalam membedakan calon santri yang Diterima dan Tidak Diterima.

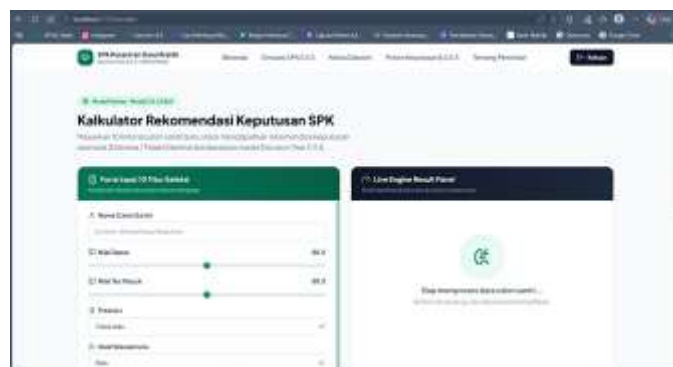
3.2 Hasil Antarmuka

Halaman Beranda dirancang sebagai gerbang utama antarmuka sistem yang memuat informasi statistik performa model, rincian kriteria penilaian seleksi, serta alur kerja algoritma C4.5.



Gambar 5. Tampilan Antarmuka Halaman Utama (Beranda / *Home Page*) Aplikasi SPK

Sistem keamanan antarmuka dibangun menggunakan arsitektur State Management berbasis React Context API (AuthContext) dan pembatas rute (*Route Guard / Protected Route*). Modul ini memastikan bahwa fitur krusial seperti kalkulator simulasi SPK dan manajemen dataset hanya dapat diakses oleh Panitia Seleksi yang terotentikasi (Fitur *Login*). Selanjutnya, halaman simulasi SPK dirancang menggunakan tata letak dua kolom responsif (*split-screen layout*). Sisi kiri memuat formulir masukan 10 kriteria kandidat santri, sedangkan sisi kanan berfungsi sebagai panel eksekusi inferensi model C4.5 secara real-time.



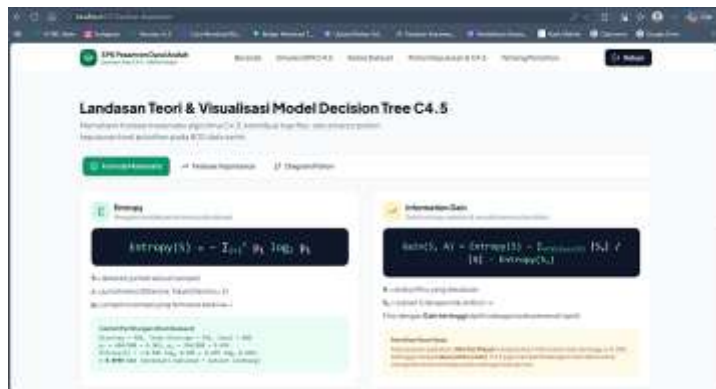
Gambar 6. Tampilan Antarmuka *Workspace* Simulasi SPK C4.5 dan Panel Hasil Klasifikasi *Real-Time*

Kemudian sistem menyediakan modul pencetakan Surat Rekomendasi Hasil Seleksi berbasis format dokumen baku Pesantren Darul Arafah Medan serta dilengkapi dengan fungsi pencarian, penyaringan data, serta file reader untuk impor dan ekspor berkas CSV.



Gambar 7. Tampilan Antarmuka Modal Pratinjau Surat Rekomendasi Hasil Seleksi Siap Cetak

Selanjutnya halaman visualisasi pohon keputusan dan landasan model C4.5 dimana halaman ini juga dikembangkan untuk menyediakan transparansi algoritma (Explainable AI), menyajikan landasan teori perhitungan matematis C4.5, peringkat pengaruh atribut (*Feature Importance*), serta diagram struktur pohon keputusan hasil pemangkasan (*pruned tree*) [20].



Gambar 8. Tampilan Antarmuka Visualisasi Pohon Keputusan dan Formulasi Metode C4.5

4. KESIMPULAN

Berdasarkan hasil penelitian, penerapan algoritma C4.5 (*Decision Tree*) berhasil digunakan untuk mengklasifikasikan penerimaan santri baru berdasarkan 10 kriteria yang mencakup aspek akademik, keagamaan, karakter, dan kedisiplinan dari 800 data historis. Model memperoleh akurasi 71%, *precision* 67%, *recall* 81%, dan *F1-score* 74%, yang menunjukkan kinerja klasifikasi yang tinggi dan relatif seimbang. Hasil *feature importance* menunjukkan bahwa Hasil_Wawancara menjadi atribut paling dominan dengan bobot 0,4215 (42,15%), diikuti oleh Nilai_Rapor sebesar 0,2640 (26,40%), sedangkan kriteria keagamaan seperti Tajwid, Hafalan_Juz, Makhraj, dan Fasohah secara kumulatif berkontribusi sebesar 27,25%. Selain itu, model berhasil diimplementasikan dalam SPK berbasis web yang mendukung pengolahan data calon santri, visualisasi aturan keputusan, serta penerbitan rekomendasi hasil seleksi, sehingga dapat membantu proses penerimaan santri secara lebih objektif, terukur, dan transparan.

REFERENCES

- [1] R. Hasan, M. A. Islam, and S. Ahmad, "Educational data mining for student performance and admission prediction: A systematic literature review," *Educ. Inf. Technol.*, vol. 28, no. 3, pp. 2931–2954, 2023.
- [2] F. Yanti, A. Fauzan, D. Apriyanti, M. A. Wahyudin, and M. Penelitian, "Data Mining untuk Klasifikasi Penerimaan Peserta Didik Baru dengan Menerapkan Algoritma Decision Tree," vol. 3, no. 4, pp. 511–515, 2025.
- [3] M. F. Roini and others, "Sistem Pendukung Keputusan Penerimaan Santri Baru untuk Menentukan Kelas Diniyah Menggunakan Metode MOORA," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 5, no. 1, pp. 195–202, 2021.
- [4] C. Chaka, "Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: An overview of reviews," *J. E-Learning Knowl. Soc.*, vol. 18, no. 2, pp. 58–69, 2022.
- [5] M. Khan and others, "Machine learning based predictive modeling of student admission in higher education institutions," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 8701–8714, 2021.
- [6] A. Wibowo and H. Kurniawan, "Penerapan Data Mining untuk Prediksi Penerimaan Peserta Didik Baru Menggunakan Algoritma Klasifikasi," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 3, pp. 555–562, 2023.
- [7] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Comput. Sci.*, vol. 2, no. 3, p. 160, 2021.
- [8] N. A. Ramadhani and H. A. Rosyid, "Review: Algoritma Data Mining untuk Klasifikasi Data," *J. Inov. Teknol. dan Edukasi Tek.*, vol. 2, no. 12, pp. 550–556, 2022, doi: 10.17977/um068v2i122022p550-556.
- [9] K. Aprilia, U. Khaira, and B. Ferdian Hutabarat, "Komparasi Algoritma Naive Bayes Dan Decision Tree Pada Klasifikasi Penerimaan Peserta Didik Baru," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 10, no. 2, pp. 547–553, 2025, doi: 10.51876/simtek.v10i2.1694.
- [10] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann Publishers, 1993.
- [11] H. H. Kusumawardani, I. Rosyadi, F. A. Artanto, F. I. Arzha, and N. A. Rachmayani, "Analisis decision tree dalam pengaruh digital marketing terhadap penerimaan siswa baru," *Remik Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 6, no. 2, pp. 225–231, 2022.
- [12] D. P. Pasaribu, T. Sutrisno, and B. Mulyawan, "Klasifikasi Hasil Belajar Siswa Menggunakan Metode C4.5 Berdasarkan Riwayat Akademik," *J. Ilmu Komput. dan Sist. Inf.*, vol. 11, no. 2, pp. 1–6, 2023.

- [13] H. Al-Shehri, M. Al-Qurishi, A. Al-Rakhami, and A. Alamri, "Application of decision tree classification algorithms for student performance prediction," *IEEE Access*, vol. 10, pp. 35454–35465, 2022.
- [14] S. A. Situmorang, "Seleksi Calon Siswa Baru dengan Metode Decision Tree," *J. Inform. dan Sist. Inf.*, vol. 5, no. 2, pp. 112–120, 2024.
- [15] A. M. Dinata, R. A. Asmara, and D. Puspitasari, "Pengaruh Pembagian Dataset Menggunakan Stratified K-Fold Cross Validation Terhadap Kinerja Algoritma Klasifikasi," *J. Teknoinfo*, vol. 18, no. 1, pp. 45–52, 2024.
- [16] X. Wang, Y. Li, and Z. Wang, "An improved C4.5 algorithm based on attribute weighting and its application in education data mining," *Appl. Sci.*, vol. 11, no. 15, p. 6813, 2021.
- [17] A. Kristina, "Penerapan Algoritma C4.5 Untuk Klasifikasi Kelayakan Penerima Bantuan Pendidikan," *J. HOAQ (High Qual. Appl. Comput.)*, vol. 13, no. 2, pp. 45–52, 2025.
- [18] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: An Overview," *arXiv Prepr. arXiv2008.05756*, pp. 1–17, 2020.
- [19] V. Y. D. Wijaya and others, "Penerapan Data Mining Dalam Prediksi Kinerja Akademik Menggunakan Algoritma Klasifikasi," *Infomedia J. Tek. Inform. dan Multimed.*, vol. 9, no. 1, pp. 45–53, 2024.
- [20] R. Confalonieri, T. Canna, T. Weyde, and V. Katos, "Historical perspective, technical roadmap, and open challenges of Explainable AI (XAI) in decision support systems," *Artif. Intell. Rev.*, vol. 54, no. 5, pp. 3775–3843, 2021.