

PREDIKSI KELAYAKAN SISWA SMA NEGERI JAKARTA SELATAN MENGUNAKAN ALGORITMA C.45 BEDASARKAN ZONASI, PRESTASI, DAN RAPOR

Attoric Hikmal Fajar ¹, Samuel Bernard Jeffersen ², Rizki Fadilla ³,
Ahmad Refi Zaini ⁴, Adi Sucipto ⁵, Baginda Oloan Lubis⁶

^{1,2,3,4,5} Sistem Informasi, Bina Sarana Informatika

⁶ Rekayasa Perangkat Lunak, Bina Sarana Informatika

Jl. Kramat Raya No. 98 Jakarta, Indonesia

attorickuliah@gmail.com

ABSTRAK

Proses seleksi siswa di SMA Negeri Jakarta Selatan menghadapi tantangan dalam memastikan penilaian kelayakan calon siswa dilakukan secara objektif dan transparan. Beberapa faktor utama, seperti zonasi, nilai rapor, prestasi akademik, dan prestasi non-akademik, menjadi indikator penting dalam menentukan penerimaan siswa baru. Namun, pendekatan manual sering mengalami kesulitan dalam mengintegrasikan berbagai kriteria tersebut dengan keterbatasan kapasitas daya tampung sekolah. Penelitian ini bertujuan untuk memprediksi kelayakan calon siswa menggunakan algoritma C4.5, yang dikenal efektif untuk mengolah data klasifikasi yang kompleks dan bervariasi. Dataset penelitian diperoleh dari data pendaftaran siswa ke SMA Negeri di wilayah Jakarta Selatan pada tahun ajaran tertentu, mencakup beragam variabel penting. Model C4.5 menghasilkan pohon keputusan yang mampu memetakan kelayakan calon siswa secara sistematis dan akurat. Pengujian menunjukkan bahwa model ini memiliki tingkat akurasi 67%, yang dianggap cukup baik untuk tahap awal penerapannya dalam proses seleksi siswa. Hasil penelitian ini diharapkan dapat mendukung sekolah dalam melaksanakan seleksi siswa secara lebih transparan, efektif, dan adil dengan mempertimbangkan semua kriteria secara menyeluruh. Selain itu, implementasi model ini diharapkan menjadi acuan bagi institusi pendidikan lain yang menghadapi tantangan serupa dalam menilai kelayakan calon siswa secara holistik.

Kata kunci : Kelayakan Siswa, Algoritma C4.5, Data Mining, Klasifikasi

1. PENDAHULUAN

Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional menyatakan bahwa pendidikan adalah upaya yang disadari dan direncanakan untuk menciptakan lingkungan belajar serta proses pembelajaran yang memungkinkan peserta didik secara aktif mengembangkan potensinya. Hal ini bertujuan agar mereka memiliki kekuatan spiritual keagamaan, kemampuan pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang dibutuhkan bagi diri mereka sendiri dan masyarakat.[1]

Pendidikan tidak hanya merupakan proses perkembangan yang sistematis dan terencana, tetapi juga upaya untuk mendorong potensi individu serta mempersiapkan mereka agar dapat berkontribusi secara optimal dalam kehidupan bermasyarakat. [2]

Pendidikan menengah atas di Indonesia, terutama di kota besar seperti Jakarta Selatan, memiliki peran strategis dalam mempersiapkan siswa untuk melanjutkan ke pendidikan tinggi atau memasuki dunia kerja. SMA Negeri menjadi pilihan favorit bagi siswa, terutama dari wilayah sekitarnya. Namun, keterbatasan kapasitas dan tingginya jumlah pendaftar mengharuskan proses seleksi penerimaan siswa baru dilakukan secara hati-hati untuk menjamin keadilan dan ketepatan. Salah satu metode yang diterapkan adalah sistem zonasi, yang bertujuan tidak hanya untuk memastikan pemerataan akses pendidikan tetapi juga

mendorong kesetaraan antara sekolah unggulan dan sekolah lainnya. Dengan sistem ini, diharapkan seluruh sekolah di Indonesia dapat menjalankan peran dan tanggung jawab yang setara, sekaligus mengurangi kesenjangan pendidikan antara sekolah di daerah perkotaan dan pedesaan. [3]

Meskipun sistem zonasi efektif dalam memastikan pemerataan akses pendidikan, hal ini belum mampu sepenuhnya menilai kelayakan calon siswa secara menyeluruh. Pelaksanaan zonasi, bersama dengan penggunaan Rencana Pelaksanaan Pembelajaran (RPP), masih dianggap memberatkan guru dan kurang efektif dalam meningkatkan kualitas pendidikan. Oleh karena itu, diperlukan kriteria tambahan seperti nilai rapor, prestasi akademik, dan prestasi non-akademik untuk menciptakan proses seleksi yang lebih objektif dan adil.[4]

Selain sistem zonasi, nilai rapor atau prestasi belajar merupakan indikator penting karena mencerminkan konsistensi akademik siswa selama menempuh pendidikan sebelumnya. Rapor atau catatan prestasi belajar meliputi hasil dari berbagai mata pelajaran, sehingga memberikan gambaran yang komprehensif tentang kompetensi akademik siswa.[5]

Prestasi belajar merujuk pada pencapaian keberhasilan seseorang dalam mencapai suatu tujuan sebagai hasil dari upaya yang telah dilakukan. [6]

Hasil dari proses pembelajaran diukur dan dinyatakan dalam bentuk angka, huruf, simbol, atau

kalimat yang menggambarkan keberhasilan siswa selama pembelajaran berlangsung. Di sisi lain, prestasi akademik mencakup pencapaian siswa dalam lomba atau kompetisi, baik di tingkat lokal, nasional, maupun internasional, yang berkaitan dengan bidang akademik tertentu. Prestasi akademik adalah istilah yang digunakan untuk menunjukkan pencapaian atau tingkat keberhasilan seseorang dalam mencapai tujuan sebagai hasil dari usaha belajar yang telah dilakukan secara optimal.[7]

Prestasi akademik sering dianggap penting karena mencerminkan kemampuan siswa dalam menghadapi persaingan yang lebih luas.[8]

Selain itu, prestasi non-akademik, seperti dalam bidang olahraga, seni, atau keterampilan lainnya, semakin dihargai dalam proses seleksi karena menunjukkan kemampuan siswa di luar lingkungan kelas serta kontribusinya dalam kegiatan yang melibatkan kreativitas, kerjasama, dan disiplin.

Menurut [9] Prestasi non-akademik merupakan gabungan dari kata "prestasi" dan "non-akademik." Seperti yang telah dijelaskan sebelumnya, prestasi adalah hasil dari suatu kegiatan yang telah diselesaikan atau diciptakan, baik secara individu maupun kelompok. Kegiatan non-akademik biasanya disebut sebagai kegiatan ekstrakurikuler, yang merujuk pada berbagai aktivitas di luar jam sekolah yang dilakukan siswa untuk mengembangkan potensi, minat, bakat, dan hobi mereka. Dengan mempertimbangkan empat faktor ini: zonasi, nilai rapor, prestasi akademik, dan prestasi non-akademik: seleksi siswa diharapkan dapat memberikan penilaian yang lebih seimbang antara aspek akademik dan non-akademik.

Untuk mengatasi berbagai faktor yang saling mempengaruhi, data mining memainkan peran penting dalam membantu analisis yang kompleks. Salah satunya adalah algoritma C4.5, yang digunakan untuk memprediksi keputusan dengan menerapkan serangkaian aturan keputusan. [10]

Pohon keputusan ini akan membagi data menjadi beberapa node berdasarkan atribut yang paling berpengaruh terhadap hasil atau keluaran yang diinginkan. Dalam konteks penelitian ini, algoritma C4.5 digunakan untuk menentukan faktor kelayakan siswa. Algoritma C4.5 berfungsi dengan mengukur information gain dari setiap atribut, yang menunjukkan seberapa efektif suatu atribut dalam memisahkan data. Atribut dengan information gain tertinggi akan menjadi akar atau node utama pada pohon keputusan, sehingga pohon yang dihasilkan dapat memberikan prediksi yang lebih akurat.

Algoritma C4.5 dipilih karena kemampuannya yang fleksibel dalam menangani data kompleks dengan berbagai jenis atribut, baik kategorikal maupun numerik, yang relevan untuk analisis faktor kelayakan siswa. Dalam penelitian ini, pohon keputusan yang dihasilkan diharapkan dapat menggambarkan hubungan antara zonasi, nilai rapor, prestasi akademik, dan prestasi non-akademik dengan status kelayakan siswa secara akurat. Dengan demikian, model ini dapat

menjadi dasar bagi sekolah untuk mengembangkan sistem penerimaan siswa baru yang lebih efisien, transparan, dan berbasis data, sehingga mendukung tercapainya tujuan seleksi yang adil dan tepat sasaran.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah proses ekstraksi informasi penting dari suatu kumpulan data. Informasi ini diperoleh melalui proses yang kompleks, yang melibatkan penggunaan kecerdasan buatan, teknik statistik, matematika, pembelajaran mesin, dan lainnya.[11]

Dalam konteks ini, data mining mencakup pengumpulan, pembersihan, dan transformasi data, serta pencarian pola tersembunyi yang dapat memberikan wawasan lebih dalam mengenai data tersebut. [12]

Dengan memanfaatkan berbagai teknik ini, data mining dapat mengidentifikasi pola, tren, atau hubungan yang sebelumnya tidak terlihat, sehingga membantu dalam pengambilan keputusan yang lebih baik dalam berbagai aplikasi, termasuk dalam analisis kelayakan siswa untuk diterima di sekolah menengah atas di kota besar seperti Jakarta Selatan. [13][14]

2.2. Algoritma C4.5

Algoritma C4.5 adalah sekumpulan algoritma yang digunakan dalam teknik klasifikasi untuk pembelajaran mesin dan penambangan data. [15]

Tujuannya adalah pembelajaran terawasi, di mana setiap entri dalam kumpulan data dijelaskan oleh sekumpulan nilai atribut, dan setiap entri tersebut termasuk dalam salah satu dari berbagai kelas yang berbeda dan saling eksklusif.[16]

Proses ini menggunakan pembelajaran terawasi, di mana data dijelaskan melalui nilai atribut dan dikelompokkan ke dalam kelas yang berbeda. Algoritma C4.5 menjadi dasar bagi pohon keputusan yang dibangun berdasarkan dataset pelatihan yang berisi entri-entri dalam basis data. [17]

Proses pembuatan pohon keputusan menggunakan C4.5 dimulai dengan memilih atribut terbaik berdasarkan rasio gain, yang mengukur sejauh mana atribut efektif dalam memisahkan data.[18]

Selanjutnya, data dibagi berdasarkan nilai atribut tersebut, dan proses ini diulang hingga seluruh data terklasifikasi. C4.5 mampu menangani atribut numerik dan kategorikal, serta data yang hilang, menjadikannya metode yang efektif untuk klasifikasi data yang kompleks.

Tahapan dalam penerapan pemodelan Algoritma C4.5 adalah sebagai berikut:

- Menyusun data pelatihan, yang diperoleh dari hasil kuesioner.
- Menentukan akar pohon keputusan yang akan dibangun dengan menghitung nilai gain tertinggi dari setiap atribut berdasarkan nilai indeks entropi terendah. Berikut adalah rumus untuk menghitung atau mencari nilai indeks entropi:

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

Keterangan :

S : Subnet Data

n : Jumlah Respondent [19]

pi : Proporsi dari Si terhadap S

- c. Selanjutnya menghitung nilai Gain, berikut rumus untuk menghitung nilai Gain:

$$Gain(S, A)$$

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

Keterangan :

S : Subnet Data

A : Atribut butir pertanyaan [19]

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

- d. Langkah kedua harus diulang terus hingga setiap cabang terpenuhi.
e. Proses pembuatan pohon keputusan berhenti ketika semua cabang dalam node N berada dalam kelas yang sama, tidak ada atribut lagi yang dapat dibagi, dan tidak ada cabang yang kosong.
f. Informasi yang dihasilkan dibagi menjadi dua kelompok, yaitu Puas dan Tidak Puas.

2.3. Decision Tree

Decision tree adalah salah satu metode klasifikasi yang menggunakan struktur pohon, di mana setiap node mewakili atribut, cabang-cabang mewakili nilai atribut, dan daun-daun mewakili kelas.[20]

Node yang terletak di bagian atas pohon keputusan disebut root. Decision tree adalah metode klasifikasi yang paling banyak digunakan. Proses pembuatan pohon keputusan dimulai dari node akar hingga node daun secara rekursif, di mana setiap percabangan menunjukkan kondisi dan setiap ujung pohon menunjukkan keputusan. [21]

Selain pembangunannya yang relatif cepat, hasil dari model ini juga mudah dipahami. Dalam decision tree, terdapat tiga jenis node, yaitu Root Node, Internal Node, dan Leaf Node. [22].

2.4. Penelitian Terdahulu

Salah satu penelitian menerapkan algoritma C4.5 untuk memprediksi prestasi siswa di SMPN 51 Jakarta. Studi ini menunjukkan bahwa algoritma C4.5 dapat digunakan untuk menentukan prediksi prestasi siswa dengan akurasi yang baik. [23]

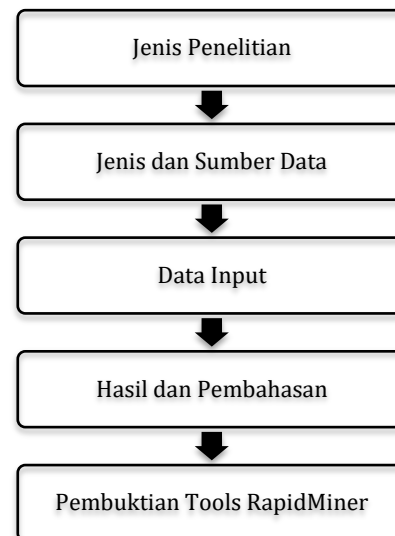
Penelitian lain mengimplementasikan algoritma C4.5 dalam sistem prediksi prestasi siswa di SMK Al-Ishlah. Dengan mempertimbangkan kriteria seperti status ekonomi, kedisiplinan, dan presensi, penelitian ini berhasil memprediksi prestasi siswa dan membantu pihak sekolah dalam pengambilan keputusan.[24]

Selain itu, studi lain membandingkan penerapan algoritma Decision Tree C4.5 dan Naïve Bayes dalam

memprediksi kelulusan siswa. Hasilnya menunjukkan bahwa algoritma C4.5 lebih akurat dalam memprediksi kelulusan siswa dibandingkan dengan Naïve Bayes. [25]

3. METODE PENELITIAN

Gambar 1 merupakan bagan metode penelitian yang menjelaskan Langkah-langkah penelitian ini.



Gambar 1. Bagan Metode Penelitian

3.1. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif analitis. Metode ini dipilih untuk menganalisis data secara sistematis dan menjelaskan hubungan antara faktor zonasi, nilai rapor, prestasi akademik, serta prestasi non-akademik terhadap kelayakan siswa. Proses analisis dilakukan menggunakan algoritma C4.5, yang memungkinkan pemetaan data menjadi pohon keputusan untuk menghasilkan klasifikasi yang akurat.

3.2. Jenis dan Sumber Data Penelitian

Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder referensi dari arsip resmi penerimaan peserta didik baru (PPDB) yang tersedia di <https://arsip.siap-ppdb.com>. Situs ini menyediakan daftar siswa yang mendaftar ke SMA Negeri di Jakarta Selatan, termasuk informasi detail seperti zonasi, nilai rapor, serta prestasi akademik dan non-akademik. Data diakses dengan menelusuri daftar siswa dan membuka detail masing-masing individu. Meskipun data tidak tersedia dalam format CSV, informasi yang dikumpulkan dari situs ini kemudian diolah secara manual dan disusun dalam format yang sesuai untuk analisis.

3.3. Data

Tabel 1 merupakan data penelitian yang akan diolah.

Tabel 1. Data Penelitian Merupakan Data Set Yang Sudah Diolah.

Jarak (km)	Prestasi Akademik	Prestasi Non Akademik	Nilai Rapor (Rata-rata)
3,45	Ada	Ada	88
10,18	Tidak Ada	Ada	82
8,12	Ada	Tidak Ada	91
8,25	Ada	Tidak Ada	85
7,3	Ada	Ada	90
5,5	Tidak Ada	Tidak Ada	70
4,2	Ada	Tidak Ada	78
12,7	Tidak Ada	Tidak Ada	72
11,4	Ada	Tidak Ada	80
6,9	Ada	Tidak Ada	79
16,4	Tidak Ada	Ada	75
17,8	Ada	Tidak Ada	82
18,9	Tidak Ada	Ada	76
14,5	Ada	Tidak Ada	84
9,4	Tidak Ada	Tidak Ada	70
3,2	Tidak Ada	Ada	74
7,1	Ada	Tidak Ada	79
15,9	Ada	Ada	85
4,7	Tidak Ada	Ada	72
20,6	Tidak Ada	Tidak Ada	70
10,9	Ada	Tidak Ada	76
18,2	Ada	Tidak Ada	78
19,7	Tidak Ada	Ada	82
8,3	Ada	Tidak Ada	80
13,6	Ada	Tidak Ada	88
3,9	Tidak Ada	Ada	75
16,2	Ada	Tidak Ada	87
12,3	Tidak Ada	Ada	70
14,1	Tidak Ada	Ada	78
17,2	Ada	Tidak Ada	91
6,5	Tidak Ada	Tidak Ada	71
5,3	Ada	Tidak Ada	77
9,6	Tidak Ada	Ada	73
19,4	Tidak Ada	Ada	80
3,7	Ada	Tidak Ada	85
13,1	Tidak Ada	Ada	74
11,1	Ada	Tidak Ada	81
10,6	Tidak Ada	Ada	78
4,4	Tidak Ada	Tidak Ada	70
5,2	Ada	Tidak Ada	88
14,8	Ada	Tidak Ada	77
15,3	Tidak Ada	Ada	73
16,7	Tidak Ada	Tidak Ada	72
13,9	Ada	Tidak Ada	79
11,6	Tidak Ada	Ada	74
20,3	Tidak Ada	Tidak Ada	70
4,6	Ada	Tidak Ada	85
17,6	Ada	Tidak Ada	80
18,3	Tidak Ada	Ada	74
7,8	Tidak Ada	Tidak Ada	70
6,4	Ada	Tidak Ada	79
9,1	Tidak Ada	Ada	73
11,7	Tidak Ada	Tidak Ada	70
14,2	Ada	Tidak Ada	86
13,2	Tidak Ada	Ada	72
16,3	Ada	Tidak Ada	84
5,1	Tidak Ada	Tidak Ada	72
10,8	Ada	Tidak Ada	90
19,3	Tidak Ada	Ada	78
7,2	Tidak Ada	Tidak Ada	70
17,1	Ada	Tidak Ada	91
6,8	Tidak Ada	Ada	76

Jarak (km)	Prestasi Akademik	Prestasi Non Akademik	Nilai Rapor (Rata-rata)
9,5	Tidak Ada	Tidak Ada	70
4,5	Ada	Ada	75
18,1	Ada	Tidak Ada	80
20,1	Tidak Ada	Ada	79
15,7	Tidak Ada	Ada	77
7,9	Tidak Ada	Ada	72
9,2	Ada	Tidak Ada	80
12,2	Tidak Ada	Ada	78
3,4	Ada	Tidak Ada	88
5,6	Tidak Ada	Tidak Ada	74
11,5	Tidak Ada	Ada	73
8,4	Ada	Tidak Ada	75
16,9	Ada	Tidak Ada	86
12,8	Tidak Ada	Ada	77
4,1	Ada	Tidak Ada	80
17,4	Tidak Ada	Ada	73
9,9	Ada	Tidak Ada	75
14,4	Tidak Ada	Ada	80
7,6	Tidak Ada	Ada	71
13,3	Ada	Tidak Ada	78
5,8	Tidak Ada	Ada	74
3,3	Ada	Tidak Ada	85
8,1	Tidak Ada	Tidak Ada	70
12,5	Tidak Ada	Ada	79
6,6	Tidak Ada	Tidak Ada	72
11,8	Ada	Tidak Ada	80
4,9	Ada	Ada	75
20,7	Tidak Ada	Tidak Ada	71
18,5	Tidak Ada	Ada	75
9,7	Tidak Ada	Tidak Ada	70
7,5	Ada	Tidak Ada	79
5,4	Tidak Ada	Tidak Ada	72
14,3	Ada	Tidak Ada	80
3,6	Ada	Tidak Ada	87
19,8	Tidak Ada	Ada	78
10,4	Ada	Tidak Ada	79
16,1	Ada	Tidak Ada	90
13,4	Tidak Ada	Tidak Ada	72
5,2	Ada	Tidak Ada	80
6,1	Tidak Ada	Tidak Ada	70
7,3	Tidak Ada	Ada	74
14,7	Ada	Tidak Ada	80
12,9	Tidak Ada	Ada	75
8,5	Tidak Ada	Tidak Ada	72
10,3	Tidak Ada	Ada	77
4,3	Ada	Tidak Ada	85
17	Tidak Ada	Ada	76
9,3	Ada	Tidak Ada	80
5,9	Tidak Ada	Tidak Ada	70
6,7	Ada	Tidak Ada	77
7,5	Tidak Ada	Tidak Ada	72

Tabel 2. Data set Sudah di Olah

No	Zonasi	Prestasi Akademik	Prestasi non Akademik	Nilai_Rapor	Kelayakan
1	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
2	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
3	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
4	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
5	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
6	Diluar_Area	Ada	Tidak_Ada	Cukup	Layak
7	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
8	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
9	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak

No	Zonasi	Prestasi Akademik	Prestasi non Akademik	Nilai_Rapor	Kelayakan
10	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
11	Diluar_Area	Ada	Ada	Tidak_Cukup	Layak
12	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
13	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
14	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
15	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
16	Diluar_Area	Ada	Tidak_Ada	Cukup	Layak
17	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
18	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
19	Diluar_Area	Tidak_Ada	Tidak_Ada	Cukup	Tidak Layak
20	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Layak
21	Diluar_Area	Ada	Tidak_Ada	Cukup	Layak
22	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
23	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
24	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
25	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
26	Diluar_Area	Ada	Ada	Tidak_Cukup	Layak
27	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
28	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
29	Diluar_Area	Ada	Tidak_Ada	Cukup	Layak
30	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
31	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
32	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
33	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
34	Diluar_Area	Ada	Ada	Tidak_Cukup	Layak
35	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
36	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
37	Diluar_Area	Ada	Tidak_Ada	Cukup	Layak
38	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
39	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
40	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
41	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
42	Diluar_Area	Ada	Ada	Tidak_Cukup	Layak
43	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
44	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
45	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
46	Diluar_Area	Tidak_Ada	Tidak_Ada	Cukup	Tidak Layak
47	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Layak
48	Diluar_Area	Ada	Tidak_Ada	Cukup	Layak
49	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
50	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
51	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
52	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
53	Diluar_Area	Ada	Ada	Tidak_Cukup	Layak
54	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
55	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
56	Diluar_Area	Ada	Tidak_Ada	Cukup	Layak
57	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
58	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
59	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
60	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
61	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
62	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
63	Diluar_Area	Ada	Tidak_Ada	Cukup	Layak
64	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
65	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
66	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
67	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
68	Diluar_Area	Ada	Ada	Tidak_Cukup	Layak
69	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
70	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
71	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
72	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak

No	Zonasi	Prestasi Akademik	Prestasi non Akademik	Nilai Rapor	Kelayakan
73	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
74	Diluar_Area	Ada	Ada	Tidak_Cukup	Layak
75	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
76	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
77	Diluar_Area	Ada	Tidak_Ada	Cukup	Layak
78	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
79	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
80	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
81	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
82	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
83	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
84	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
85	Diluar_Area	Ada	Ada	Tidak_Cukup	Layak
86	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
87	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
88	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
89	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
90	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
91	Diluar_Area	Ada	Ada	Tidak_Cukup	Layak
92	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Layak
93	Diluar_Area	Ada	Tidak_Ada	Cukup	Layak
94	Diluar_Area	Tidak_Ada	Ada	Tidak_Cukup	Tidak Layak
95	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
96	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak
97	Diluar_Area	Ada	Tidak_Ada	Tidak_Cukup	Tidak Layak
98	Diluar_Area	Ada	Ada	Tidak_Cukup	Layak
99	Diluar_Area	Tidak_Ada	Ada	Cukup	Layak
100	Didalam_Area	Tidak_Ada	Tidak_Ada	Tidak_Cukup	Layak

3.4. Menghitung Gain dan Entropy

Menentukan atribut sebagai akar dan menghitung nilai Gain dari atribut tersebut. Untuk memilih atribut sebagai akar, ini didasarkan pada nilai gain tertinggi

dari atribut-atribut yang ada. Nilai Entropy diperlukan untuk untuk menentukan nilai Gain tertinggi.

Menurut perhitungan yang ditunjukkan pada tabel dibawah ini, atribut dengan nilai gain tertinggi adalah 0.122525248.

Tabel 3. Hasil Perhitungan Gain dan Entropy

		Jumlah (S)	Layak	Tidak Layak	Entropy	Gain
TOTAL		100	70	30	0,8812909	
Zonasi						0,07725667
	Didalam_Area	27	25	2	0,38094659	
	Diluar_Area	73	45	28	0,96051871	
Prestasi Akademik						0,02988366
	Ada	35	20	15	0,98522814	
	Tidak Ada	65	50	15	0,77934984	
Prestasi Non Akademik						0,01095497
	Ada	46	35	11	0,79359123	
	Tidak Ada	54	35	19	0,93571105	
Nilai Rapor						0,12252525
	Cukup	34	32	2	0,32275696	
	Tidak Cukup	66	38	28	0,98337619	

4. HASIL DAN PEMBAHASAN

Hasil penelitian ini diharapkan untuk menentukan nilai akurasi Algoritma C4.5 dalam memprediksi kelayakan siswa SMA Negeri Jakarta Selatan berdasarkan zonasi, prestasi, dan rapor. Penelitian ini dilakukan dengan mengolah data, memilih atribut yang relevan, serta menguji data tersebut melalui perhitungan manual dan perangkat lunak RapidMiner. Hasil dari penelitian ini berupa proses perhitungan berdasarkan algoritma C4.5 yang

mendukung pengambilan keputusan secara lebih efektif dan akurat.

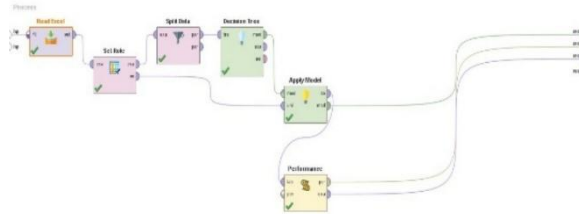
4.1. Pengujian algoritma C4.5

Pembuatan algoritma C4.5 dilakukan pada kumpulan dataset, dengan data yang sudah dijelaskan sebelumnya pengujian data dapat dihitung dengan beberapa langkah-langkah perhitungan Entropy dan Gain pada setiap atribut yang dimiliki diterima yang sebagai variabel target. Dengan menghitung nilai gain di setiap atribut, lalu dengan membandingkan nilai

gain tersebut yang nilainya tertinggi akan digunakan sebagai simpul akar atau atribut akar.

4.2. Pengujian Data Dengan RapidMiner

Pada proses pengujian data yang sudah di import, peneliti menggunakan data training dan data testing untuk menentukan sebuah model decision tree dengan menggunakan perbandingan 65% untuk data training dan 35% testing.



Gambar 2. Pengujian Data Dengan Rapid Miner

4.3. Hasil Pemodelan Pohon Keputusan



Gambar 3. Pohon Keputusan

Pohon keputusan C4.5 yang terlihat pada gambar dapat dibaca dari atas ke bawah, dimulai dari akar (simpul pertama paling atas) hingga ke daun (simpul terluar yang tidak memiliki cabang). Berikut adalah cara membaca pohon keputusan berdasarkan simpul-simpulnya:

- Jika Zonasi berada di Dalam Area, maka siswa dinyatakan Layak.
- Jika diluar area maka dilihat berdasarkan nilai rapor jika cukup dinyatakan layak
- Jika tidak cukup melihat berdasarkan prestasi non akademik
- Jika tidak ada maka tidak layak
- Jika ada melihat prestasi akademik, jika ada prestasi akademik dinyatakan layak.
- Jika tidak dinyatakan tidak layak

Pohon keputusan ini menghasilkan aturan-aturan klasifikasi berdasarkan kombinasi atribut yang ada, sehingga memudahkan dalam menentukan kelayakan siswa sesuai kriteria yang ditentukan.

4.4. Hasil Accuracy

	True Layak	True Tidak Layak	Overall Accuracy
pred Layak	87	4	94,37%
pred Tidak Layak	3	26	89,66%
Overall	85,71%	88,57%	

Gambar 4. Hasil Accuracy

Hasil pengukuran akurasi data yang diperoleh dari data training dapat dilihat pada gambar di atas dengan nilai accuracy mencapai 93%. Dari tabel dapat diketahui :

- Prediksi siswa Layak
 - Jumlah data true Layak yang diprediksi sebagai Layak adalah 70 siswa.
 - Tidak ada data true Tidak yang salah diprediksi sebagai Layak.
 - Precision untuk prediksi kelas Layak adalah 94,37%.
- Prediksi siswa Tidak Layak
 - Jumlah data true Tidak yang diprediksi sebagai Tidak Layak adalah 30 siswa.
 - Tidak ada data true Layak yang salah diprediksi sebagai Tidak Layak.
 - Precision untuk prediksi kelas Tidak Layak adalah 89,66%.
- Class Recall
 - Untuk kelas Layak, nilai recall mencapai 95,71%.
 - Untuk kelas Tidak Layak, nilai recall mencapai 86,67%.

Hasil ini menunjukkan model memiliki performa yang sempurna dalam memprediksi data siswa berdasarkan atribut yang digunakan.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil memanfaatkan algoritma C4.5 untuk memprediksi kelayakan siswa SMA Negeri Jakarta Selatan berdasarkan faktor zonasi, nilai rapor, prestasi akademik, dan prestasi non-akademik. Hasil analisis menunjukkan bahwa zonasi menjadi faktor dominan dalam menentukan kelayakan siswa, diikuti oleh nilai rapor sebagai indikator akademik utama. Model pohon keputusan yang dihasilkan memiliki akurasi sebesar 93%, yang menunjukkan kemampuan moderat dalam memprediksi kelayakan siswa. Untuk meningkatkan akurasi model dan penerapannya yang lebih luas, beberapa langkah dapat dipertimbangkan. Pertama, penambahan variabel seperti kondisi sosial ekonomi, jarak rumah ke sekolah, atau preferensi siswa dapat memperkaya model. Selain itu, algoritma lain seperti Random Forest atau Gradient Boosting dapat diuji untuk meningkatkan akurasi prediksi.

Peningkatan kualitas data melalui pembersihan dan normalisasi juga diperlukan untuk mengurangi noise dalam dataset. Validasi model dapat ditingkatkan menggunakan metode seperti k-fold cross-validation untuk memastikan keandalan dan generalisasi hasil prediksi. Pendekatan hybrid yang

menggabungkan data mining dengan metode statistik atau simulasi kebijakan juga dapat memberikan wawasan lebih komprehensif. Uji implementasi pada beberapa sekolah dengan karakteristik berbeda penting dilakukan untuk mengevaluasi adaptabilitas model. Terakhir, kolaborasi dengan pihak sekolah, pemerintah, dan masyarakat diperlukan untuk memastikan model selaras dengan kebijakan pendidikan dan kebutuhan lokal. Dengan menerapkan langkah-langkah ini, model dapat menjadi alat pendukung keputusan yang lebih efektif untuk kebijakan zonasi dan proses seleksi siswa yang lebih objektif, efisien, dan transparan.

DAFTAR PUSTAKA

- [1] S. Ujud, T. D. Nur, Y. Yusuf, N. Saibi, and M. R. Ramli, "Penerapan Model Pembelajaran Discovery Learning Untuk Meningkatkan Hasil Belajar Siswa Sma Negeri 10 Kota Ternate Kelas X Pada Materi Pencemaran Lingkungan," *J. Bioedukasi*, vol. 6, no. 2, pp. 337–347, 2023, doi: 10.33387/bioedu.v6i2.7305.
- [2] N. S. Anggraini Dewi, "Menuju Pendidikan Berkelanjutan : Implementasi Tujuan Pembangunan Berkelanjutan (Sdgs) Dalam Mewujudkan," *J. Penelit. Pendidik. Indones.*, vol. 1, no. 3, pp. 189–197, 2024.
- [3] Nurdin, H. S. A. Zulfa, A. F. B. Setiawan, Nurkhaerattunnisa, and R. M. Fadli, "Kebijakan Zonasi Sekolah Di Indonesia : Tinjauan Literatur Tentang Implementasi Dan Dampaknya Terhadap Mutu Pendidikan," *J. MEDIA Akad.*, vol. 2, no. 10, pp. 1–16, 2024.
- [4] Y. M. Daud, "Perkembangan Kebijakan Pendidikan di Indonesia: A Systematic Literature Review," *Intelektualita*, vol. 13, no. 1, pp. 115–131, 2024, doi: 10.22373/ji.v13i1.24871.
- [5] M. Khairat and M. G. Adiyanti, "Self-esteem dan prestasi akademik sebagai prediktor pubjective well-being remaja awal," *J. Psychol.*, vol. 1, no. 3, pp. 180–191, 2015.
- [6] T. Supariyadi, I. Mahfud, and R. Marsheilla Aguss, "Hubungan Tingkat Kebugaran Jasmani Terhadap Prestasi Belajar Penjas Tahun 2021," *J. Arts Educ.*, vol. 2, no. 2, pp. 60–71, 2022, doi: 10.33365/jae.v2i2.109.
- [7] T. M. S. Manurung, "Pengaruh Motivasi dan Perilaku Belajar Terhadap Prestasi Akademik Mahasiswa," *JAS-PT J. Anal. Sist. Pendidik. Tinggi*, vol. 1, no. 1, p. 17, 2017, doi: 10.36339/jaspt.v1i1.36.
- [8] B. O. Lubis, B. Santoso, R. T. Yunandar, A. Salim, and D. Oscar, "Implementasi Aplikasi Raport Digital Berbasis Website dengan Metode Global Extreme Programming," *J. Teknol. Inform. dan Komput.*, vol. 9, no. 1, pp. 293–305, 2023, doi: 10.37012/jtik.v9i1.1394.
- [9] Nurul Alifah, "Analisis Manajemen Kesiswaan dalam Meningkatkan Prestasi Akademik dan Non Akademik di SMP Negeri 1 Banyuglugur Situbondo," *J. Ilm. Manaj. dan Kewirausahaan*, vol. 3, no. 3, pp. 75–90, 2023, doi: 10.55606/jimak.v3i3.2142.
- [10] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 2, p. 126, 2016, doi: 10.35314/isi.v1i2.131.
- [11] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 1, pp. 13–21, 2021, doi: 10.30813/jbase.v4i1.2729.
- [12] C. I. Ifaru *et al.*, "Penentuan Sparepart Genset Paling Sering Digunakan Pada Operator Indosat Ooredoo Hutchison Dengan Algoritma C4.5," *J. Teknol. Sist. Inf.*, vol. 5, no. 1, pp. 300–309, 2024, doi: 10.35957/jtsi.v5i1.7481.
- [13] R. Risda, F. J. Septriwinti, and F. Nasution, "Pendekatan Pemrosesan Informasi," *MUDABBIR J. Reserch Educ. Stud.*, vol. 3, no. 1, pp. 49–59, 2023, doi: 10.56832/mudabbir.v3i1.260.
- [14] S. Dwiasnati, W. Gunawan, R. R. Oprasto, B. O. Lubis, and B. Santoso, *Algoritma dan Pemrograman Implementasi Menggunakan Python*. Bandar Lampung: CV. Keranjang Teknologi Media, 2023.
- [15] D. M. Musa *et al.*, "Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4.5," *J. Teknol. Inform. dan Komput.*, vol. 10, no. 1, pp. 168–182, 2024, doi: 10.37012/jtik.v10i1.1985.
- [16] A. Purwanto, "Jurnal Teknoinfo," *Tong Sampah Pint. Dengan Perintah Suara Guna Menghilangkan Perilaku Siswa Membuang Sampah Sembarangan Di Sekol.*, vol. 14, pp. 48–58, 2020.
- [17] A. Nugraha, I. Maulana, A. Susilo, Y. Irawan, and A. Voutama, "(STUDI KASUS : SMK PGRI CIKAMPEK)," vol. 8, no. 5, pp. 10078–10086, 2024.
- [18] D. Jatmika and G. Taufik, "Penerapan Algoritma C4.5 Untuk Klasifikasi Keberhasilan Pengiriman Barang," *Invotek Polbeng*, vol. 8, no. 1, pp. 12–26, 2021.
- [19] F. Handayanna, "Penerapan Algoritma C4.5 Untuk Memprediksi Keluhan Pelanggan Pada Apartemen Jakarta Application of the C4.5 Algorithm to Predict Customer Complaints at Jakarta Apartments," vol. 21, no. 4, pp. 868–877, 2022.
- [20] L. Bachtiar and M. Mahradianur, "Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai," *J. Inform.*, vol. 10, no. 1, pp. 28–36, 2023, doi: 10.31294/inf.v10i1.15115.
- [21] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51,

- 2021, doi: 10.35329/jiik.v7i2.203.
- [22] S. Saifullah, M. Zarlis, Z. Zakaria, and R. W. Sembiring, "Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data," *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 1, no. 2, p. 180, 2017, doi: 10.30645/j-sakti.v1i2.41.
- [23] M. H. Sukri and Y. Handrianto, "Penerapan Algoritma C4. 5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta," *Informatics Comput. Eng. J.*, vol. 4, no. 1, pp. 11–24, 2024.
- [24] E. Martantoh and A. P. Haryanto, "Implementasi Algoritma C4.5 Untuk Sistem Prediksi Prestasi Siswa Di SMK Al-Islah Berbasis Web," *J. Teknol. Sist. Inf.*, vol. 3, no. 2, pp. 131–142, 2022, doi: 10.35957/jtsi.v3i2.2773.
- [25] Y. Susanti, M. G. Choyyin, A. Priyatna, and S. Lestari, "Perbandingan Penerapan Algoritma Decision Tree C.45 dan Naïve Bayes Dalam Analisa Kelulusan Siswa pada SMK Swadhipa 2 Natar Kabupaten Lampung Selatan," *J. SIMADA (Sistem Inf. dan Manaj. Basis Data)*, vol. 6, no. 2, pp. 117–123, 2023, doi: 10.30873/simada.v6i2.3772.