

Penerapan Algoritma C4.5 Dalam Sistem Prediksi Kelulusan Santri Tahfidz Qur'an

Putri Aprilia Siregar*, Raissa Amanda Putri

Sains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Deli Serdang
Jl. Lapangan Golf, Desa Durian Jangak, Kec. Pancur Batu Kab. Deli Serdang, Sumatera Utara, Indonesia

Email: ^{1*} putriapriliasiregar10@gmail.com, ² raissa.ap@uinsu.ac.id

Email Penulis Korespondensi: putriapriliasiregar10@gmail.com

Submitted: 19/07/2024; Accepted: 26/07/2024; Published: 27/07/2024

Abstrak—Pesantren Tahfidz Ulumul Qur'an Medan adalah lembaga pendidikan keagamaan islam yang fokus kepada pendidikan penghafal Al-Qur'an. Namun, tidak semua santri dapat lulus tepat waktu sesuai masa belajar yang ditempuh sehingga mengakibatkan penumpukan santri yang lulus tidak tepat waktu. Penentuan kelulusan santri berdasarkan beberapa kriteria yang harus dilalui oleh santri selama menempuh pendidikan di Pesantren Tahfidz Ulumul Qur'an. Oleh karena itu, Algoritma C4.5 merupakan salah satu metode dalam pembelajaran mesin yang digunakan untuk membangun pohon keputusan berdasarkan data santri sebanyak 160. Data yang digunakan dalam penelitian ini meliputi berbagai faktor yang mempengaruhi kelulusan santri, seperti Makhoriul Huruf, Hafalan, Kedisiplinan, dan Lafadz yang relevan. Hasil dari penelitian ini menunjukkan bahwa algoritma C4.5 mampu memberikan prediksi yang akurat mengenai kelulusan santri Tahfidz Qur'an dengan tingkat akurasi sebanyak 98.21%. Dimana berdasarkan hasil pengolahan data santri didapati 107 data santri yang lulus tepat waktu dan 53 tidak lulus tepat waktu.

Kata Kunci: Prediksi; Santri Tahfidz Qur'an; Algoritma C4.5; Pohon Keputusan

Abstract—Tahfidz Ulumul Qur'an Islamic Boarding School Medan is an Islamic religious educational institution that focuses on educating memorizers of the Al-Qur'an. However, not all students can graduate on time according to the study period taken, resulting in a buildup of students who graduate not on time. The determination of student graduation is based on several criteria that students must pass while studying at the Tahfidz Ulumul Qur'an Islamic Boarding School. Therefore, the C4.5 algorithm is a machine learning method used to build a decision tree based on 160 students' data. The data used in this research includes various factors that influence students' graduation, such as Makhoriul Letters, Memorizing, Discipline, and Relevant lafadz. The results of this research show that the C4.5 algorithm is able to provide accurate predictions regarding the graduation of Tahfidz Qur'an students with an accuracy rate of 98.21%. Where, based on the results of processing student data, it was found that 107 data of students graduated on time and 53 did not graduate on time.

Keywords: Prediction; Tahfidz Qur'an Students; C4.5 Algorithm; Decision Tree

1. PENDAHULUAN

Al-Qur'an adalah mukjizat terbesar dan teragung yang diberikan kepada manusia mulia, Baginda Rasulullah SAW. Di dalamnya terdapat petunjuk dan keagungan yang luar biasa, sehingga tak ada seorang pun yang mampu menandinginya sebagai pedoman meraih kemenangan dalam hidup. Sebaliknya, bagi mereka yang meremehkan atau tidak beriman kepadanya, akan mendapatkan kerugian di dunia dan akhirat. Selain itu, Al-Qur'an juga merupakan kitab suci umat Islam yang berfungsi sebagai petunjuk dan pedoman hidup bagi orang-orang yang bertakwa. Proses tahfidz Al-Qur'an pertama kali terjadi ketika wahyu pertama turun kepada Nabi Muhammad SAW di Gua Hira, kemudian Rasulullah SAW membacakan wahyu tersebut dari hafalannya kepada istrinya, Siti Khadijah. Maka, kebiasaan menghafal Al-Qur'an telah ada sejak Al-Qur'an diturunkan kepada Rasulullah SAW dan kemudian dihafal oleh para sahabat [1][2]. Tahfidz Al-Qur'an adalah metode mempelajari Al-Qur'an dengan cara menghafalnya, sedangkan tahsin Al-Qur'an bertujuan untuk memperbaiki bacaan Al-Qur'an [3]. Proses untuk hafal Al-Qur'an dianggap sebagai tantangan yang kompleks, meskipun ayat-ayatnya mudah dihafal, mempertahankan hafalan tersebut terasa sulit dalam praktiknya [4]. Pesantren di berbagai daerah telah banyak mengembangkan lembaga pendidikan sesuai perkembangan zaman namun tidak tercabut dengan akar sosial masyarakat dan tradisi keilmuannya [5].

Pesantren Tahfidz Ulumul Qur'an adalah lembaga pendidikan keagamaan islam yang fokus kepada pendidikan penghafal al-Qur'an. Penentuan kelulusan santri didasarkan kepada seberapa baik hafalan yang dimiliki oleh santri selama menempuh pembelajaran di Pesantren Tahfidz Ulumul Qur'an Medan. Data yang digunakan dalam penelitian ini meliputi berbagai faktor yang mempengaruhi kelulusan santri, seperti Makhoriul Huruf, Hafalan, Kedisiplinan, dan Lafadz yang relevan. Oleh karena itu, yang menjadi dasar bagi penulis untuk melakukan penelitian berjudul "Penerapan Algoritma C4.5 Dalam Sistem Prediksi Kelulusan Santri Tahfidz Qur'an Menggunakan Algoritma C4.5 (Studi Kasus Pesantren Ulumul Qur'an)" Penelitian ini diharapkan dapat memudahkan dalam memperoleh informasi untuk memprediksi kelulusan santri tepat waktu. Penelitian ini menggunakan beberapa kriteria seperti Makharijul Huruf, Hafalan, Kedisiplinan, dan Lafadz sebagai faktor penilaian, sehingga dengan sistem ini dapat membantu pihak pesantren dalam proses penilaian dan menghasilkan nilai yang tepat dan akurat. Peneliti menggunakan data santri Tahfidz Ulumul Qur'an pada semester ganjil tahun 2023 sebanyak 160 data.



Prediksi didefinisikan sebagai serangkaian kegiatan untuk memperkirakan sesuatu melalui urutan proses tertentu, baik terkait dengan objek maupun peristiwa yang mungkin terjadi pada masa kini dan masa mendatang, berdasarkan informasi valid yang telah diperoleh di masa lalu [6]. Prediksi ini dilakukan dengan tujuan agar apa yang akan dilakukan di masa kini dan mendatang tidak mengalami kesalahan, atau jika terjadi kesalahan, dapat diminimalisir dan diantisipasi sebelumnya [7][8]. Prediksi seringkali tidak akan persis sama dengan realitas yang terjadi, namun dengan menjalankan proses prediksi yang akurat, diharapkan dapat memberikan jawaban yang mendekati kenyataan [9].

Data mining dalam istilah sederhana adalah penemuan pola yang berguna dalam pengolahan data, data mining juga disebut sebagai ilmu pengetahuan, machine learning, dan analisis prediksi [10]. Data mining adalah proses pencarian pola atau informasi menarik pada data terpilih dengan menggunakan teknik atau metode tertentu [11]. Data mining merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi –informasi yang nantinya dapat digunakan [12]. Data mining dapat digunakan dalam penggunaan matematika, statistik, AI, machine learning dalam mengidentifikasi suatu informasi data yang besar sehingga mendapatkan suatu pengetahuan baru [13]. Sebuah solusi yang disebut RapidMiner untuk menganalisis data mining, text mining, dan analisis prediktif. Sistem aplikasi ini menggunakan berbagai teknik deskriptif dan prediktif untuk memberikan informasi kepada pengguna sehingga dapat mengambil keputusan terbaik [14].

Decision Tree adalah salah satu metode dalam Knowledge Discovery in Database (KDD). Metode ini menghasilkan pengetahuan yang diperoleh dari pengolahan database. [15]. Database akan diekstraksi untuk menghasilkan informasi atau pengetahuan yang bermanfaat. Tahapan yang dilakukan dalam KDD mencakup data mining. Data mining adalah suatu algoritma yang digunakan untuk menggali informasi berharga yang tersembunyi dalam koleksi data (database) yang sangat besar, sehingga pola-pola menarik yang sebelumnya tidak diketahui dapat ditemukan. Salah satu algoritma yang digunakan dalam data mining adalah algoritma C4.5, yang berfungsi untuk membentuk pohon keputusan dan aturan-aturan (rules) [16]. Algoritma C4.5 memetakan nilai atribut menjadi kelas yang dapat diimplementasikan untuk klasifikasi baru [17]. Algoritma ini mempunyai input berupa training samples dan samples. Training samples adalah berupa data contoh yang digunakan untuk membangun sebuah tree yang telah diuji kebenarannya [18]. Sedangkan samples merupakan field- field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data [19].

Algoritma C4.5 memiliki ide dasar yaitu pembentukan pohon keputusan (Decision Tree). Dimana Pohon keputusan termasuk metode klasifikasi dan prediksi yang terkenal karena sangat mudah dipahami dengan bahasa alami sehingga dapat diinterpretasikan dengan cepat. Pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki gain tertinggi berdasarkan nilai entropy atribut tersebut sebagai sumbu atribut klasifikasi [15]. C4.5 merupakan pembaharuan dari algoritma sebelumnya yaitu algoritma Iterative Dichotomiser3 (ID3) yang ditemukan oleh Ross Quinlan. Selanjutnya Quinlan mepresentasikan metode C4.5, dimana untuk pemilihan split atribut pada ID3 menggunakan Information Gain maka C4.5 menggunakan Gain Ratio (GR). Proses seleksi atribut terbaik ialah atribut yang memungkinkan mendapatkan ukuran pohon keputusan terkecil. Atau atribut yang dapat memisahkan objek berdasarkan kelas. Secara heuristik, atribut yang diseleksi adalah atribut yang menghasilkan node termurni (cleanest) [18]. Pengukuran kinerja yang dilakukan menggunakan sekelompok data uji untuk mengetahui prosentase precision, recall dan accuracy, menunjukkan bahwa algoritma C4.5 memiliki tingkat akurasi yang lebih tinggi dari pada algoritma lain. [20]

RapidMiner adalah perangkat lunak sumber terbuka. RapidMiner adalah solusi untuk menganalisis data mining, text mining, dan analisis prediktif. RapidMiner menggunakan berbagai teknik deskriptif dan prediktif untuk memberikan wawasan kepada pengguna sehingga mereka dapat membuat keputusan terbaik. RapidMiner memiliki sekitar 500 operator penambahan data, termasuk operator input, output, pemrosesan data, dan visualisasi. RapidMiner ditulis dalam Java, sehingga dapat berfungsi di semua sistem operasi. [21] Digunakan dalam aplikasi perusahaan dan komersial, serta penelitian, pelatihan, pendidikan, pembuatan prototipe cepat, dan pengembangan aplikasi, ini mendukung semua langkah pembelajaran mesin, termasuk persiapan data, visualisasi hasil, validasi, dan pengoptimalan [22].

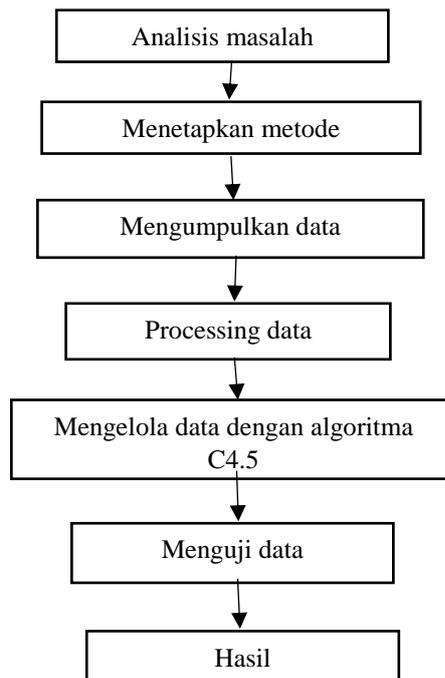
Penelitian yang dilakukan oleh [17] yang membahas tentang Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5 dikarenakan banyaknya perguruan tinggi swasta yang ada, maka persaingan antar perguruan tinggi untuk menarik minat calon mahasiswa baru sangatlah ketat. Maka untuk dapat bersaing dengan perguruan tinggi tersebut, pihak kampus perlu melakukan prediksi terhadap minat calon mahasiswa baru dengan mengetahui faktor apa yang menjadi pendorong calon mahasiswa baru memilih perguruan tinggi. Tujuan penelitian ini adalah untuk memprediksi minat calon mahasiswa dalam memilih perguruan tinggi. Dalam penelitian ini, data yang diolah adalah hasil survei yang dilakukan pada calon mahasiswa baru pada Program Studi D3-Manajemen Informatika. Terdapat 25 data sampel sebagai training dataset dengan 5 data penilaian. Data tersebut akan diolah menggunakan proses Data Mining dengan Metode Klasifikasi. Selanjutnyadengan algoritma C4.5 pada metode Klasifikasidiperoleh iterasi sampai kepada node 4 dengan 3 data penilaian yang menjadi faktor dalam penentuan minat calon mahasiswa terhadap program studi.Selanjutnya hasil pengolahan datamenggunakan metode klasifikasi dengan algoritma C4.5 tersebut, dilakukan pengujian dengan menggunakan Software Rapid Miner, dimana diperoleh tingkat akurasi 100% dengan 120 testing dataset. Hasil akhir penggunaan Data Mining dengan algoritma C4.5 ini telah mampu melakukan prediksi terhadap minat calon

mahasiswa baru berdasarkan faktor penilaian dalam memilih perguruan tinggi. Kemudian terdapat penelitian yang dilakukan oleh [23] yang membahas tentang Penerapan Algoritma C4.5 Dalam Mengukur Kepuasan Pengunjung Terhadap Fasilitas di Taman Margasatwa Jakarta. Banyak hal yang harus diperhatikan dalam mengembangkan suatu objek wisata yaitu dengan memberikan sesuatu yang sesuai dengan yang dibutuhkan dan keinginan calon pengunjung. Berdasarkan fenomena tersebut peneliti menemukan bahwa jumlah pengunjung yang terus meningkat masih ada masalah yang dihadapi di Taman Margasatwa yaitu dari segi fasilitas wisata, Penelitian ini bertujuan untuk memberi rekomendasi kepuasan pengunjung taman taman margasatwa dengan menggunakan teknik datamining penerapan datamining C4.5. Sumber data diperoleh dari penyebaran kuesioner kepada pengunjung yang pernah berkunjung ke taman margasatwa kepada 35 orang. Atribut yang digunakan sebagai parameter penilaian kepuasan pengunjung antara lain: Paling Diminati, Fasilitas Semuanya, Tempat parkir, Keamanan, dan Kebersihan. Hasil pengolahan metode C4.5 dengan menggunakan Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Berdasarkan penelitian diatas tersebut maka penelitian ini dibuat dengan perbedaan yaitu memiliki data sebanyak 160 data santri kemudian memberikan landasan yang kuat dalam membuat klasifikasi dalam menentukan prediksi kelulusan.

2. METODOLOGI PENELITIAN

2.1 Kerangka Penelitian

Penelitian ini menggunakan metode kualitatif yang merupakan metode fokus pada pengamatan yang mendalam sehingga menghasilkan suatu kejadian yang lebih komprehensif. Jenis metode penelitian deskriptif kualitatif ini memanfaatkan data kualitatif dan dijabarkan secara deskripsi. Metode deskriptif merupakan pendekatan studi objektif yang bertujuan untuk memberikan gambaran jelas mengenai suatu keadaan dengan mengumpulkan informasi dan menganalisis data. Berikut kerangka penelitian Gambar 1:



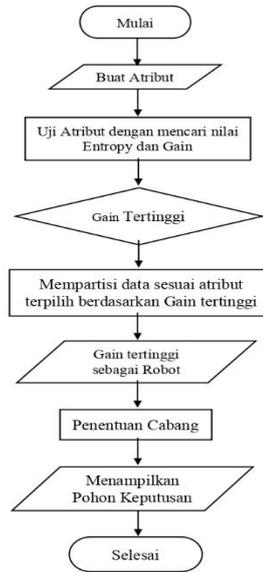
Gambar 1. Kerangka Penelitian

Berikut ini adalah penjelasan dari kerangka penelitian diatas:

1. Analisis Masalah. Tahap masalah merupakan tahap awal penelitian. Peneliti menemukan masalah berdasarkan permasalahan yang ditemukan pada data santri tahfidz Qur'an.
2. Menetapkan Metode. Untuk menemukan prediksi yang akurat maka peneliti menetapkan metode menggunakan algoritma C4.5
3. Mengumpulkan Data. Peneliti melakukan mengumpulkan data santri secara langsung (observasi).
4. Processing Data. Proses pengumpulan data mentah dan mentransformasi data tersebut menjadi informasi yang lebih akurat dan bermakna.
5. Mengelola data dengan Algoritma C4.5. Pengolahan data santri menggunakan rumus algoritma C4.5
6. Menguji Data. Peneliti menggunakan aplikasi RapidMiner yang mendukung algoritma C4.5 dan menghasilkan pohon keputusan.
7. Hasil. Tahapan ini adalah tahapan terakhir yaitu dengan adanya hasil output Lulus Tepat Waktu dan Lulus Tidak Tepat Waktu.

2.2 Rencana Penelitian

Penelitian ini menggunakan metode Algoritma C4.5 untuk melihat lulus dan tidak lulus dalam Sistem Prediksi Tahfidz Ulumul Qur’an. Peneliti menggunakan data santri Tahfidz Ulumul Qur’an tahun 2023 sebanyak 160 data. Data ini diproses dari pesantren tahfidz ulumul qur’an kota medan. Dimana atribut yang digunakan yaitu atribut makhori jul huruf, kedisiplinan, hafalan, dan tahfidz. Adapun tahapan yang dilakukan penelitian ini seperti gambar 2 sebagai berikut:



Gambar 2. Tahapan Prediksi Algoritma C4.5

Pada tahap persiapan awal, atribut yang akan digunakan ditentukan terlebih dahulu, kemudian atribut-atribut tersebut diuji dengan mencari nilai Gain tertinggi berdasarkan perhitungan entropy dari masing-masing atribut. Jika ditemukan gain tertinggi, maka nilai tersebut akan menjadi root awal. Selanjutnya, cabang ditentukan dengan cara yang sama, yaitu dengan melihat gain tertinggi dari setiap hasil partisi. Adapun rumus algoritma C4.5 sebagai berikut :

Rumus Entropy :

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

Ket : p_i = probabilitas dari setiap nilai unik pada variabel target.

Rumus Gain :

$$\text{Gain}(S, A) = E(S) - \sum_{v \in \text{Values}(A)} \left(\frac{|S_v|}{|S|} \times E(S_v) \right) \tag{2}$$

Ket: S_v = subset dari S untuk mana atribut A memiliki nilai v

3. HASIL DAN PEMBAHASAN

3.1 Processing Data

Berdasarkan data santri di pesantren tahfidz ulumul Qur’an yang telah melalui proses pembersihan, dilakukan transformasi data. Atribut yang dapat ditransformasikan meliputi Makhori jul Huruf, Hafalan, Kedisiplinan, dan Lafadz. Data yang diinput digunakan untuk merancang model Decision Tree. Transformasi ini dilakukan dengan mengklasifikasikan masing-masing atribut yang telah ditentukan. Berikut adalah tabel 1 klasifikasi untuk atribut-atribut tersebut:

Tabel 1. Klasifikasi Data Pada Atribut

No	Atribut	Class
1.	Makhori jul Huruf	50 - 70 (Kurang Lancar)
		71 - 100 (Lancar)
2.	Hafalan	50 - 74 (Cukup Sedikit)
		75 - 100 (Banyak)
3.	Kedisiplinan	45 - 70 (Kurang Disiplin)
		71 - 100 (Disiplin)



No	Atribut	Class
4.	Lafadz	50 - 65 (Cukup Mampu) 66 - 100 (Mampu)

Pada tabel 1 telah dijelaskan bahwa 4 atribut yang biasa digunakan untuk menginput data, mempunyai klasifikasi masing-masing agar mempermudah proses data mining. Ke 4 atribut tersebut akan dianalisa menggunakan algoritma C4.5.

Tabel 2. Atribut Target

No	Atribut Target	Keterangan
1.	Hasil	45 - 79 (Tidak Tepat Waktu) 80 - 100 (Tepat Waktu)

Dari tabel 2 telah dijelaskan bahwasanya mempunyai 2 target yang peneliti inginkan adalah Tidak Tepat Waktu dan Tepat Waktu. Analisis yang ditetapkan dengan 2 target data terlihat pada tabel 3 sebagai berikut:

Tabel 3. Sampel Data

No	MakhoriJul Huruf	Hapalan	Kedisiplinan	Lafadz	Hasil
1	Kurang Lancar	Banyak	Disiplin	Mampu	Tidak Tepat Waktu
2	Kurang Lancar	Cukup Sedikit	Kurang Disiplin	Mampu	Tidak Tepat Waktu
3	Kurang Lancar	Cukup Sedikit	Kurang Disiplin	Mampu	Tidak Tepat Waktu
4	Lancar	Banyak	Disiplin	Mampu	Tepat Waktu
5	Lancar	Banyak	Disiplin	Mampu	Tepat Waktu
....
....
158	Kurang Lancar	Cukup Sedikit	Kurang Disiplin	Mampu	Tidak Tepat Waktu
159	Lancar	Banyak	Disiplin	Mampu	Tepat Waktu
160	Lancar	Banyak	Disiplin	Mampu	Tepat Waktu

3.2 Mengelola Data Dengan Algoritma C4.5

Proses yang dilalui pada penelitian ini yaitu mencari nilai entropy total atau keseluruhan terlebih dahulu, kemudian mencari nilai entropy setiap atribut. Setelah nilai entropy telah didapatkan maka tahap berikutnya yaitu mencari nilai gain setiap atribut, tahap ini terus dilakukan berulang kali sampai semua atribut dilakukan atau dengan membentuk beberapa node pohon keputusan.

Perhitungan Entropy dan Gain

Menghitung Entropy Total :

$$(-107/160) * \log_2(107/160) + (-53/160) * \log_2(53/160) = 0,91619$$

Menghitung Entropy MakhoriJul Huruf

Entropy Lancar :

$$(-104/110) * \log_2(104/110) + (-6/110) * \log_2(6/110) = 0,30540$$

Entropy Tidak Lancar :

$$(-3/50) * \log_2(3/50) + (-47/50) * \log_2(47/50) = 0,32744$$

Menghitung Entropy Hapalan

Entropy Banyak :

$$(-106/114) * \log_2(106/114) + (-8/114) * \log_2(8/114) = 0,36657$$

Entropy Cukup Sedikit :

$$(-1/46) * \log_2(1/46) + (-45/46) * \log_2(45/46) = 0,15109$$

Menghitung Entropy Kedisiplinan

Entropy Disiplin :

$$(-105/116) * \log_2(105/116) + (-11/116) * \log_2(11/160) = 0,45238$$

Entropy Kurang Disiplin :

$$(-2/44) * \log_2(2/44) + (-42/44) * \log_2(42/44) = 0,26676$$

Menghitung Entropy Lafadz

Entropy Mampu :

$$(-106/138) * \log_2(106/138) + (-32/138) * \log_2(32/138) = 0,78128$$

Entropy Cukup Mampu :

$$(-1/22) * \log_2(1/22) + (-21/22) * \log_2(21/22) = 0,26676$$

Menghitung Gain MakhoriJul Huruf :

$$(0,91619) - (110/160) * 0,30540 - (50/160) * 0,32744 = 0,60390$$

Menghitung Gain Hapalan :

$$(0,91619) - (114/160) * 0,36657 - (46/160) * 0,15109 = 0,61157$$

Menghitung Gain Kedisiplinan :



$$(0,91619) - (116/160) * 0,45238) - (44/160) * 0,26676) = 0,51486$$

Menghitung Gain Lafadz :

$$(0,91619) - (138/160) * 0,78128) - (22/160) * 0,26676) = 0,20566$$

Tabel 4. Perhitungan Node 1

Node	Keterangan	Total	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain
Total		160	107	53	0,916198397	
MakhoriJul Huruf	Lancar	110	104	6	0,305400571	0,603909
	Kurang Lancar	50	3	47	0,327444919	
Hapalan	Banyak	114	106	8	0,366578013	0,611571
	Cukup Sedikit	46	1	45	0,151096971	
Kedisiplinan	Disiplin	116	105	11	0,452381623	0,514861
	Kurang Disiplin	44	2	42	0,266764988	
	Lafadz					
Mampu	Mampu	138	106	32	0,781281209	0,205663
	Cukup Mampu	22	1	21	0,266764988	

Berdasarkan perhitungan pada tabel 4 dapat diketahui bahwa atribut dengan Gain tertinggi adalah Hafalan dengan nilai 0.611571, sehingga atribut Hafalan dapat menjadi Node Akar. Pada atribut yang dijadikan akar terdapat 2 nilai yaitu “Banyak” dan “Sedikit”. Dari Nilai atribut tersebut, nilai “Lancar” mempunyai hasil “Tepat Waktu”, Sedangkan “Kurang Lancar” Mempunyai hasil “Tidak Tepat Waktu”. Jadi dapat disimpulkan bahwa nilai information gain tertinggi untuk dijadikan sebagai akar pohon keputusan.

Tabel 5. Perhitungan Node 1.2

Node	Keterangan	Total	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain
Total		114	106	8	0,366578013	
MakhoriJul Huruf	Lancar	104	103	1	0,078232465	0,217901825
	Kurang Lancar	10	3	7	0,881290899	
Kedisiplinan	Disiplin	107	104	3	0,184451863	0,14045351
	Kurang Disiplin	7	2	5	0,863120569	
Lafadz	Mampu	107	105	2	0,134029796	0,204447332
	Cukup Mampu	7	1	6	0,591672779	

Pada perhitungan tabel 5 dapat diketahui bahwa atribut dengan Gain tertinggi selanjutnya adalah MakhoriJul Huruf dengan nilai 0.217901825, sehingga atribut MakhoriJul Huruf dapat menjadi Node berikutnya. Maka Node akar selanjutnya adalah atribut Kedisiplinan yang dilanjutkan dengan Node akar terakhir adalah Lafadz.



Tabel 6. Perhitungan Node 1.3

Node	Keterangan	Tot al	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain
i.iii Makhori ul Huruf (Kurang Lancar)	Total	50	3	47	0,3274449 19	0,1175582 47
	Kedisiplin an					
	Disiplin	14	3	11	0,7495952 57	
	Kurang Disiplin	36	0	36	0	
	Lafadz					0,0373769 08
	Mampu	33	3	30	0,4394969 87	
	Cukup Mampu	17	0	17	0	

Tabel 6 di atas menunjukkan analisis pengukuran entropi dan gain. Data total terdiri dari 50 sampel, di mana 3 tepat waktu dan 47 tidak tepat waktu. Tabel ini dibagi berdasarkan dua atribut: "Kedisiplinan" dan "Lafadz." Pada atribut "Kedisiplinan," ada 14 sampel yang disiplin (3 tepat waktu, 11 tidak tepat waktu) dengan entropi 0,7459 serta 36 sampel yang kurang disiplin (semuanya tidak tepat waktu) dengan entropi 0. Dan didapatkan hasil Gain 0,1176. Pada atribut "Lafadz," ada 33 sampel yang mampu (3 tepat waktu, 30 tidak tepat waktu) dengan entropi 0,4394 serta 17 sampel yang cukup mampu (semuanya tidak tepat waktu) dengan entropi 0, kemudian didapatkan nilai Gainnya 0,0374. Data ini digunakan untuk menentukan atribut mana yang lebih berpengaruh dalam ketepatan waktu, dan didapatkan atribut "Kedisiplinan" memiliki gain tertinggi.

Tabel 7. Perhitungan Node 1.4

Node	Keterangan	Tota l	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain
i.iv Kedisiplina n (Disiplin)	Total	116	105	11	0,4523816 23	0,1275286 4
	Lafad z					
	Mampu	111	104	7	0,339486	
	Kurang Mampu	0	0	0	0	

Lanjutan tabel 7 di atas menunjukkan analisis entropi dan gain untuk atribut "Lafadz" dalam konteks "Kedisiplinan (Disiplin)" pada total 116 sampel. Dari 116 sampel tersebut, 105 tepat waktu dan 11 tidak tepat waktu, menghasilkan entropi sebesar 0,4523. Atribut "Lafadz" dibagi menjadi dua kategori: "Mampu" dan "Kurang Mampu." Sebanyak 111 sampel dikategorikan sebagai "Mampu" (104 tepat waktu, 7 tidak tepat waktu) dengan entropi 0,3395. Tidak ada sampel yang termasuk dalam kategori "Kurang Mampu." Gain yang dihasilkan dari pemisahan ini adalah 0,1275, menunjukkan bahwa atribut "Lafadz" cukup berpengaruh dalam ketepatan waktu pada kelompok yang disiplin.

Setelah seluruh atribut sudah ditemukan nilai Entropy dan Gain, lalu kita menghitung Entropy dan Gain Atribut Hafalan "Cukup Sedikit". Berikut tabel 7 :

Tabel 8. Perhitungan Node 1.2 Lanjutan

Node	Keterangan	Tot al	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain
i.ii Hapal an (Cuku p Sediki t)	Total	46	1	45	0,1510969 71	0,0663114 37
	Makhori jul Huruf					
	Lancar	6	1	5	0,6500224 22	
	Kurang Lancar	40	0	40	0	

Node	Keterangan	Total	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain
Kedisiplinan	Disiplin	9	1	8	0,503258335	0,052633383
	Kurang Disiplin	37	0	37	0	
Lafadz	Mampu	31	1	30	0,205592508	0,012545498
	Cukup Mampu	15	0	15	0	

Berdasarkan hasil perhitungan pada Tabel 8 diatas, didapatkan bahwa faktor Makhori jul Huruf memiliki gain tertinggi sebesar 0,066311437. Pada kategori ini, individu yang "Lancar" dalam Makhori jul Huruf memiliki entropy sebesar 0,650022422, sedangkan yang "Kurang Lancar" memiliki entropy sebesar 0. Hal ini menunjukkan bahwa kelancaran dalam Makhori jul Huruf merupakan faktor yang signifikan dalam menentukan ketepatan waktu. Setelah memilih faktor dengan gain terbesar, yaitu Makhori jul Huruf, analisis dilanjutkan dengan perhitungan atribut lainnya untuk kategori yang "Lancar".

Atribut Kedisiplinan dan Lafadz dianalisis kembali. Faktor Kedisiplinan memiliki gain sebesar 0,052633383, dengan individu yang "Disiplin" memiliki entropy sebesar 0,503258335 dan yang "Kurang Disiplin" memiliki entropy sebesar 0. Ini menunjukkan bahwa disiplin juga berpengaruh terhadap ketepatan waktu dalam kategori yang "Lancar" dalam Makhori jul Huruf. Selanjutnya, faktor Lafadz memiliki gain terendah sebesar 0,012545498, dengan individu yang "Mampu" memiliki entropy sebesar 0,205592508 dan yang "Cukup Mampu" memiliki entropy sebesar 0. Dari analisis ini, dapat disimpulkan bahwa meskipun semua faktor memiliki pengaruh, Makhori jul Huruf adalah faktor yang paling signifikan dalam mempengaruhi ketepatan waktu, diikuti oleh Kedisiplinan dan Lafadz, dan analisis atribut dilakukan secara berkelanjutan untuk setiap kategori yang diidentifikasi.

Tabel 8. Perhitungan Node 1.3 Lanjutan

Node	Keterangan	Total	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain	
i.iii Makhori jul Huruf (Lancar)	Total	110	104	6	0,305400571	0,246398525	
	Kedisiplinan	Disiplin	102	102	0		0
		Kurang Disiplin	8	2	6		0,811278124
	Lafadz	Mampu	105	103	2		0,136059286
		Cukup Mampu	5	1	4		0,721928095

Setelah didapatkan Gain Atribut Makhori jul Huruf kategori Lancar, dilakukan lagi perhitungan untuk atribut Kedisiplinan dan Lafadz pada sub-kategori tersebut. Hasil perhitungan menunjukkan bahwa faktor Kedisiplinan memiliki gain tertinggi sebesar 0,246398525, dengan individu yang "Disiplin" memiliki entropy sebesar 0 dan yang "Kurang Disiplin" memiliki entropy sebesar 0,811278124. Faktor Lafadz memiliki gain sebesar 0,142710884, dengan individu yang "Mampu" memiliki entropy sebesar 0,136059286 dan yang "Cukup Mampu" memiliki entropy sebesar 0,721928095. Hal ini menunjukkan bahwa kedisiplinan adalah faktor paling signifikan dalam kategori Makhori jul Huruf Lancar, diikuti oleh Lafadz, sehingga analisis atribut dilakukan secara berkelanjutan untuk setiap kategori yang diidentifikasi.

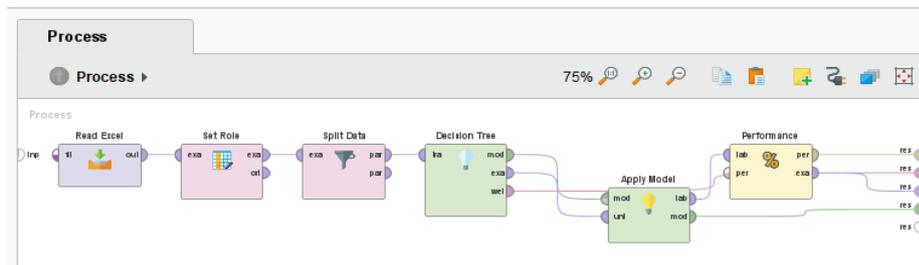
Tabel 9. Perhitungan Node 1.4 Lanjutan

Node	Keterangan	Tota l	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain
i.iv Kedisiplinan (Kurang Disiplin)	Total	44	2	42	0,266764988	0,03300231
	Lafadz					
	Mampu	27	2	25	0,380946586	
	Cukup Mampu	17	0	17	0	

Berdasarkan tabel 9 di atas, analisis dilakukan untuk menentukan pengaruh Lafadz terhadap ketepatan waktu pada kategori "Kurang Disiplin" dengan total 44 data, di mana 2 data tepat waktu dan 42 data tidak tepat waktu, menghasilkan entropy sebesar 0,266764988. Faktor Lafadz memiliki gain sebesar 0,03300231, menunjukkan pengaruh yang rendah terhadap ketepatan waktu dalam kategori ini. Pada sub-kategori Lafadz, individu yang "Mampu" terdiri dari 27 data, dengan 2 data tepat waktu dan 25 data tidak tepat waktu, menghasilkan entropy sebesar 0,380946586. Sementara itu, individu yang "Cukup Mampu" terdiri dari 17 data, dengan semua data tidak tepat waktu, sehingga entropy-nya adalah 0. Hal ini menunjukkan bahwa kemampuan lafadz sedikit berpengaruh dalam kategori kurang disiplin, namun tetap penting untuk analisis..

3.3 Proses Pengujian Data Menggunakan Rapidminer

Setelah data dianalisis dan diklasifikasikan menggunakan metode Algoritma C4.5 maka untuk tahap selanjutnya adalah dengan pembuktian dari analisis perhitungan manual tersebut. Adapun aplikasi yang digunakan dalam pengujian prediksi lulus tepat waktu dan tidak tepat waktu ini adalah menggunakan aplikasi rapidminer. Pada tahap pengujian model ini data digunakan telah melewati processing. Adapun desain model proses pembentukan Pohon Keputusan dari Algoritma C4.5 pada gambar 3 sebagai berikut :



Gambar 3. Desain Proses Pohon Keputusan

Pada Gambar 3 dapat dijelaskan sebagai berikut :

1. Read Excel berfungsi untuk mengimport data set dengan ekstensi .xlsx
2. Set Role berfungsi untuk menentukan atribut yang menjadi label kelas
3. Setel Parameter untuk membagi data menjadi dua bagian yaitu data training (70%) dan data testing (30%).
4. Lalu memasukkan Train Decision Tree Model digunakan untuk penelitian ini
5. Setelah itu apply model to test data
6. Evaluate model performance iala yang digunakan untuk mengukur performance akurasi dari model.

3.4 Hasil Pengujian Data

Setelah proses desain dijalankan maka akan didapatkan hasil pohon keputusan algoritma c4.5 seperti Gambar 4 dibawah ini:



Gambar 4. Tampilan Hasil Decision Tree

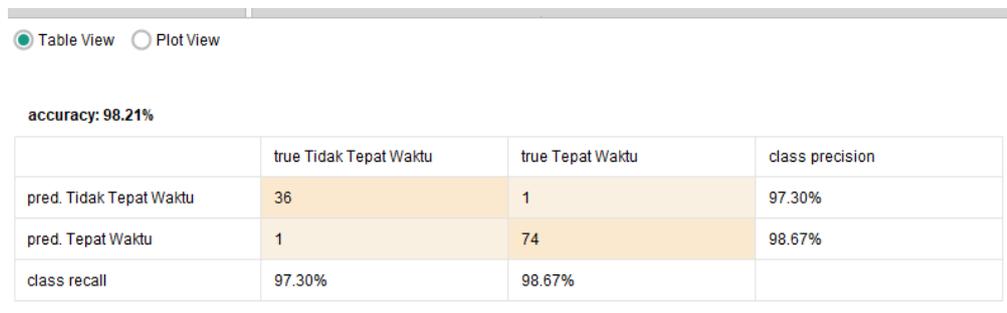
Tree

```

MakhoriJul Huruf = Kurang Lancar
| Hapalan = Banyak
| | Lafadz = Cukup Mampu: Tidak Tepat Waktu {Tidak Tepat Waktu=5, Tepat Waktu=0}
| | Lafadz = Mampu: Tepat Waktu {Tidak Tepat Waktu=1, Tepat Waktu=2}
| Hapalan = Cukup Sedikit: Tidak Tepat Waktu {Tidak Tepat Waktu=28, Tepat Waktu=0}
MakhoriJul Huruf = Lancar
| Kedisiplinan = Disiplin: Tepat Waktu {Tidak Tepat Waktu=0, Tepat Waktu=72}
| Kedisiplinan = Kurang Disiplin
| | Hapalan = Banyak: Tidak Tepat Waktu {Tidak Tepat Waktu=1, Tepat Waktu=1}
| | Hapalan = Cukup Sedikit: Tidak Tepat Waktu {Tidak Tepat Waktu=2, Tepat Waktu=0}
    
```

Gambar 5. Hasil Rule Decision Tree

Pada penelitian yang didukung hasil pemrosesan data menggunakan software RapidMiner didapat hasil akurasi penerapan algoritma C4.5 yaitu 98,21%. Untuk melihat nilai akurasi dapat dilihat pada gambar 6 sebagai berikut :



accuracy: 98.21%

	true Tidak Tepat Waktu	true Tepat Waktu	class precision
pred. Tidak Tepat Waktu	36	1	97.30%
pred. Tepat Waktu	1	74	98.67%
class recall	97.30%	98.67%	

Gambar 6. Nilai Akurasi Algoritma C4.5

Dengan pengolahan data menggunakan aplikasi RapidMiner menghasilkan nilai akurasi sistem sebanyak 98,21%. Dari gambar dijelaskan bahwa prediksi tidak tepat waktu adalah 36 dan prediksi tepat waktu adalah 1 dengan nilai precision sebesar 97,30% dan nilai recall sebesar 98,67%.

4. KESIMPULAN

Dari hasil pembahasan tentang prediksi kelulusan santri tahfidz Qur'an menggunakan Algoritma C4.5, dapat disimpulkan bahwa dari 160 data santri dengan secara observasi, 107 di antaranya lulus tepat waktu dengan kriteria MakhoriJul huruf yang lancar, hafalan yang banyak, disiplin, dan kemampuan lafadz yang baik. Sementara itu, terdapat 53 santri yang lulus tidak tepat waktu karena tidak memenuhi kriteria tersebut. Kriteria sebagai faktor penilaian, sehingga dengan sistem ini dapat membantu pihak pesantren dalam proses penilaian dan menghasilkan nilai yang tepat dan akurat. Tingkat akurasi data dengan menggunakan RapidMiner pada 160 data santri ini adalah 98,21%.

REFERENCES

- [1] M. N. 3 Nailil Maghfiroh 1, Saadi 2, “Pengembangan sistem evaluasi pendidikan islam pondok pesantren di indonesia,” *Turots J. Pendidik. Islam di Indones.*, vol. 5, no. 2, pp. 35–46, 2023.
- [2] R. L. S. Aini and A. R. Afendi, “Implementasi Metode Syafa’ati dalam Pembelajaran Tahfidz di Pondok Pesantren Darul Ikhlas Marangkayu,” *J. Educ.*, vol. 5, no. 4, pp. 15759–15770, 2023, [Online]. Available: [https://repository.uinsi.ac.id/handle/123456789/2791%0Ahttps://repository.uinsi.ac.id/bitstream/handle/123456789/2791/Implementasi Metode Syafa’ati dalam Pembelajaran Tahfidz di Pondok Pesantren Darul Ikhlas Marangkayu.pdf?sequence=1&isAllowed=y](https://repository.uinsi.ac.id/handle/123456789/2791%0Ahttps://repository.uinsi.ac.id/bitstream/handle/123456789/2791/Implementasi%20Metode%20Syafa%27ati%20dalam%20Pembelajaran%20Tahfidz%20di%20Pondok%20Pesantren%20Darul%20Ikhlas%20Marangkayu.pdf?sequence=1&isAllowed=y)
- [3] A. Azwar, T. Tampubolon, M. H. Riduan, and A. R. Matondang, “Penerapan Program Tahfidz dan Tahsin Al-Qur’an di Pondok Pesantren Darul Qur’an Kabupaten Deli Serdang,” *J. Rev. Pendidik. dan Pengajaran*, vol. 7, no. 3, pp. 9419–9423, 2024.
- [4] M. Y. Zulfikar, Hafidz, and S. Azzahro, “Penerapan Metode Talaqqi dalam Program Tahfidz Anak Usia Dini di Rumah Tahfidz Desa Beji,” *Didakt. J. Kependidikan*, vol. 13, no. 2, pp. 1755–1766, 2024, doi: 10.58230/27454312.589.
- [5] M. Marhamah, A. Lutfhi, N. Nahuda, and M. H. Rasyid, “Penyuluhan Edukatif ‘ Penguatan Nilai Karakter Bagi Pembentukan Kepribadian Di Pondok Pesantren Tahfidz Mazro’atul Lughoh Pare Kediri Jawa Timur,”” *J. PkM (Pengabdian Kpd. Masyarakat)*, vol. 6, no. 5, p. 516, 2023, doi: 10.30998/jurnalpkm.v6i5.19536.
- [6] I. Ubaedi and Y. M. Djaksana, “Optimasi Algoritma C4.5 menggunakan Metode Forward Selection dan Stratified Sampling untuk Prediksi Kelayakan Kredit,” *J. Sist. Inf.*, vol. 9, no. 1, pp. 17–26, 2022.
- [7] F. Riandari and A. Simangunsong, “Penerapan Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa,” *J. Mantik Penusa*, vol. 3, no. 2, pp. 1–7, 2019.
- [8] S. Ucha Putri, E. Irawan, F. Rizky, S. Tunas Bangsa, P. A. -Indonesia Jln Sudirman Blok No, and S. Utara, “Implementasi



- Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5,” J. Ilm. Matrik, vol. 2, no. 1, pp. 39–46, 2021.
- [9] N. Y. Lumban Gaol, “Aplikasi Prediksi Penjualan AC Menggunakan Decision Tree Dengan Algoritma C4.5,” J. Inf. Teknol., vol. 2, no. 1, pp. 23–29, 2020.
- [10] R. R. Andarista and A. Jananto, “Penerapan Data Mining Algoritma C4. 5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor,” J. Tekno Kompak, vol. 16, no. 2, pp. 29–43, 2022, [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/1525%0Ahttps://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/download/1525/944>
- [11] C. R. Aditya Nugroho and T. Kristiana, “Penerapan Algoritma C4.5 Untuk Kepuasan Pelanggan Toko Online Parfume Chantik,” J. Algoritma, vol. 3, no. 1, pp. 10–21, 2022, doi: 10.35957/algoritme.v3i1.3169.
- [12] F. Pirmansyah and T. Wahyudi, “Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Evaluasi Anggota Satuan Pengamanan Studi Kasus Pt. Yimm Pulogadung,” J. Indones. Manaj. Inform. dan Komun., vol. 4, no. 3, pp. 1566–1580, 2023, doi: 10.35870/jimik.v4i3.370.
- [13] K. Handoko, P. Simanjuntak, E. Hutabri, and E. Erlin, “Penerapan algoritma c4.5 untuk penentuan jurusan siswa sekolah menengah atas,” J. TEKINKOM, vol. 6, no. 1, pp. 153–157, 2023, doi: 10.37600/tekinkom.v6i1.788.
- [14] I. Zulkarnaen, O. E. Wulandari, Padulah, and H. Kurnia, “Identifikasi pohon keputusan hipertensi dengan sistem RapidMiner dan metode klasifikasi,” Trop. Public Heal. J., vol. 3, no. 2, pp. 63–71, 2023, doi: 10.32734/trophico.v3i2.13226.
- [15] N. A. Banyal and S. Surianti, “Implementasi Data Mining Untuk Prediksi Penyakit Malaria Tropika Dengan Algoritma C4.5,” J. Ilm. matrik, vol. 25, no. 3, 2023.
- [16] W. Yusnaeni and Widiarina, “Penerapan Algoritma C4.5 Dalam Prediksi Resiko Diabetes Tahap Awal (Early Stage Diabetes),” J. Tek. Comput. AMIK BSI, vol. 8, no. 1, pp. 56–60, 2022, doi: 10.31294/jtk.v4i2.
- [17] S. U. Putri, E. Irawan, and F. Rizky, “Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5,” J. Penerapan Sist. Inf. (Komputer Manajemen), vol. 2, no. 1, pp. 39–46, 2021, doi: 10.34010/jamika.v11i2.5506.
- [18] M. S. Hasibuan and S. Suhardi, “Analisis Sentimen Kebijakan Vaksin Covid-19 Menggunakan SVM dan C4.5,” J. Tek. Elektro Dan Komput. TRIAC, vol. 3, no. 2, pp. 19–21, 2022.
- [19] R. L. Septia Aini and A. R. Afendi, “Implementasi Metode Syafa’ati dalam Pembelajaran Tahfidz di Pondok Pesantren Darul Ikhlas Marangkayu,” J. Educ., vol. 5, no. 4, pp. 15759–15770, 2020, doi: 10.37034/jidt.v2i1.22.
- [20] N. Nurmala, E. S. Susanto, and I. M. Widiarta, “Implementasi Data Mining untuk memprediksi Kecocokan Gaya Belajar Siswa Sekolah Dasar Menggunakan Algoritma C4.5,” J. Rekayasa Inf. Swadharma, vol. 4, no. 2, pp. 1–10, 2024.
- [21] N. Manullang, R. W. Sembiring, I. Gunawan, I. Parlina, and I. Irawan, “Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5,” J. Ilmu Komput. dan Teknol., vol. 2, no. 2, pp. 1–5, 2021, doi: 10.35960/ikomti.v2i2.700.
- [22] D. Nisrina, T. Informatika, and U. N. Waluyo, “Analisis Kepuasan Konsumen Menggunakan Metode Algoritma C4 . 5 Berbasis Rapidminer pada PT . Adeaksa,” J. Means (Media Inf. Anal. dan Sist., vol. 9, no. 1, pp. 26–33, 2024.
- [23] A. Izyuddin and S. Wibisono, “Aplikasi Prediksi Penjualan Ac Menggunakan Decision Tree dengan Algoritma C4.5,” J. Manaj. Inform. Sist. Inf., vol. 3, no. 2, pp. 146–156, 2020, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/infortech73>