

Implementasi Algoritma C4.5 Dalam Memprediksi Tingkat Kelulusan Peserta Didik Kelas IX Pada SMP Negeri 184 Jakarta

Marveline Putri¹, Yana Iqbal Maulana²

Universitas Bina Sarana Informatika¹²

e-mail: marveline12ginting@gmail.com¹, yana.yim@bsi.ac.id²

Abstrak - Kelulusan dari sekolah menengah pertama memegang peran yang sangat penting bagi peserta didik, karena itu akan menjadi langkah awal yang menentukan untuk melanjutkan ke jenjang pendidikan selanjutnya, yaitu sekolah menengah atas. Dalam penelitian ini, fokus utamanya adalah untuk memprediksi tingkat kelulusan peserta didik di SMP Negeri 184 Jakarta. Data nilai kelulusan yang dikumpulkan dari peserta didik kelas IX pada tahun ajaran 2022/2023 digunakan sebagai dataset untuk dilakukan pengolahan menggunakan metode data mining dengan algoritma klasifikasi C4.5. Implementasi algoritma ini dilakukan melalui aplikasi RapidMiner, yang telah terbukti efektif dalam memprediksi tingkat kelulusan peserta didik di SMP Negeri 184 Jakarta. Pemilihan metode klasifikasi algoritma C4.5 didasarkan pada faktor kemudahan implementasi dan tingkat akurasi yang dapat diterima dalam melakukan prediksi. Hasil dari implementasi algoritma C4.5 berdasarkan data dari 355 peserta didik menunjukkan bahwa terdapat 224 peserta yang diprediksi lulus dan 123 peserta yang diprediksi tidak lulus, dengan tingkat akurasi mencapai 93.54%. Dengan adanya prediksi tingkat kelulusan yang akurat, diharapkan dapat memberikan informasi yang berharga bagi pihak sekolah dalam mengambil keputusan dan memberikan bimbingan kepada peserta didik untuk mencapai kelulusan yang sukses.

Kata Kunci : Sekolah, Kelulusan, Data Mining, Algoritma C4.5, RapidMiner

Abstract - The graduation from junior high school plays a crucial role for students as it serves as a determining step to continue their education to the next level, which is senior high school. The main focus of this research is to predict the graduation rate of students at SMP Negeri 184 Jakarta. The collected graduation score data from the 9th-grade students in the academic year 2022/2023 is utilized as the dataset and processed using the data mining method with the C4.5 classification algorithm. The implementation of this algorithm is performed through the RapidMiner application, which has proven to be effective in predicting the graduation rate of students at SMP Negeri 184 Jakarta. The selection of the C4.5 classification algorithm is based on factors such as ease of implementation and acceptable accuracy in prediction. The results of implementing the C4.5 algorithm based on the data from 355 students indicate that 222 students are predicted to graduate, while 123 students are predicted to not graduate, with an accuracy rate of 93.54%. With the availability of accurate graduation rate predictions, it is expected to provide valuable information for the school authorities to make decisions and provide guidance to students in achieving successful graduation.

Keywords: School, Graduation, Data Mining, C4.5 Algorithm, RapidMiner.

I. PENDAHULUAN

Kelulusan peserta didik adalah sebuah indikator penting yang mencerminkan keberhasilan pembelajaran dan persiapan mereka untuk jenjang pendidikan berikutnya. Standar kelulusan yang harus dipenuhi oleh peserta didik mencakup penyelesaian seluruh mata pelajaran, nilai sikap yang baik, partisipasi dan kelulusan ujian sekolah, dan tingkat kehadiran yang memadai. Nilai akademik juga menjadi faktor penentu dalam kelulusan, mempengaruhi apakah peserta didik dapat melanjutkan ke jenjang pendidikan lebih tinggi atau harus mengulang kelas. Kelulusan peserta didik memiliki dampak yang signifikan, baik bagi pihak sekolah maupun peserta didik itu sendiri.

Bagi pihak sekolah, kelulusan peserta didik dapat mempengaruhi status akreditasi sekolah. Kelulusan yang tinggi menunjukkan bahwa sekolah berhasil memberikan pendidikan yang memadai dan efektif kepada peserta didik. Hal

ini dapat meningkatkan reputasi dan citra sekolah di masyarakat, serta meningkatkan kepercayaan orang tua dan masyarakat terhadap kualitas pendidikan yang diberikan.

Bagi peserta didik, kelulusan memiliki dampak yang besar terhadap masa depan mereka. Kelulusan adalah prasyarat penting untuk melanjutkan ke jenjang pendidikan yang lebih tinggi, seperti perguruan tinggi atau universitas. Dengan meraih kelulusan, peserta didik memiliki kesempatan untuk mengembangkan pengetahuan, keterampilan, dan potensi diri mereka melalui pendidikan lanjutan. Kelulusan juga dapat membuka peluang kerja yang lebih baik di masa depan.

Prediksi tingkat kelulusan dapat membantu memahami potensi kelulusan peserta didik berdasarkan data nilai yang ada. Dalam konteks ini, data mining menjadi metode yang relevan untuk menggali informasi berharga dari kumpulan data, memungkinkan pembentukan

pola dan hubungan yang dapat mendukung pengambilan keputusan. Data mining adalah kumpulan proses yang digunakan untuk mengeksplorasi guna menemukan nilai dalam bentuk informasi juga relasi-relasi kompleks yang terpendam dari basis data (Utomo & Purba, 2019). Data mining merupakan kumpulan prosedur atau metode yang berguna untuk menganalisis nilai tambah dari sekelompok data berupa informasi yang belum ditemukan terdahulu (Azhar dkk., 2022).

Data mining memiliki teknik, metode, atau algoritma yang beragam. Salah satu teknik yang merupakan bagian penting dalam data mining yaitu teknik klasifikasi (*classification*). Klasifikasi adalah metode yang digunakan untuk mempelajari data set untuk memperoleh hubungan antar data yang membentuk suatu pattern sehingga dapat diperoleh knowledge atau pengetahuan (Etriyanti dkk., 2020).

Sementara algoritma C4.5 adalah salah satu pendekatan yang efektif dalam menganalisis data nilai peserta didik dan memprediksi kelulusan. Algoritma ini dikembangkan oleh Ross Quinlan pada tahun 1993 sebagai pengembangan dari algoritma sebelumnya yang dikenal dengan nama ID3 (*Iterative Dichotomiser 3*). Tujuan dari algoritma C4.5 adalah untuk membangun sebuah pohon keputusan yang dapat digunakan untuk mengklasifikasikan data baru berdasarkan atribut-atribut yang ada.

Algoritma C4.5 adalah metode data mining klasifikasi dan prediksi yang populer yang digunakan untuk membentuk decision tree (pohon keputusan). Salah satu keunggulan utama algoritma C4.5 adalah kemampuan untuk membuat pohon keputusan (*decision tree*) yang efisien untuk mengatur atribut tipe (diskrit dan numerik). Selain itu, algoritma ini mudah ditafsirkan dan memiliki tingkat akurasi yang dapat diterima (Muhamad dkk., 2019).

Penelitian ini berfokus pada penerapan algoritma C4.5 melalui perangkat lunak RapidMiner untuk memprediksi tingkat kelulusan peserta didik di SMP Negeri 184 Jakarta. RapidMiner adalah *software* atau perangkat lunak independen yang berguna sebagai mesin data mining untuk menganalisis data yang mampu digabungkan terhadap produknya sendiri (Maulid, 2023). RapidMiner merupakan perangkat lunak independen yang diterapkan untuk menganalisis data dan mesin penambangan data, yang dapat diintegrasikan dengan bahasa-bahasa pemrograman secara mudah (Prasetyo et al., 2021).

Data nilai peserta didik kelas IX diintegrasikan ke dalam algoritma C4.5, membentuk pohon keputusan yang akan memberikan hasil prediksi kelulusan. Penelitian ini memandang bahwa penerapan algoritma C4.5 melalui RapidMiner akan memudahkan

prediksi tingkat kelulusan dengan tingkat akurasi yang memadai.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang berharga bagi pengambil kebijakan pendidikan dalam meningkatkan efektivitas kurikulum dan mendukung peserta didik menuju kelulusan yang sukses.

II. METODOLOGI PENELITIAN

Dalam penelitian ini tahapan yang dilakukan adalah meliputi:

1. Pengumpulan Data

Data yang diperlukan sebagai sampel penelitian dalam memprediksi tingkat kelulusan peserta didik kelas IX diperoleh dari catatan nilai peserta didik di SMP Negeri 184 Jakarta selama tahun ajaran 2022-2023. Sampel yang akan digunakan sebagai data training berjumlah 355 peserta didik, yang secara representatif mencakup populasi siswa kelas IX. Data terdiri dari 4 atribut (indikator penilaian) dan 1 atribut hasil. Atribut yang digunakan dalam pengujian algoritma C4.5 meliputi nama siswa, nilai karakter, nilai rata-rata rapor, jumlah her, dan status kelulusan (lulus atau tidak lulus).

2. Pembentukan Model Algoritma C4.5

Pada penelitian ini menggunakan metode data mining dengan model Algoritma C4.5 yang merupakan salah satu algoritma pembelajaran mesin yang digunakan untuk membangun pohon keputusan (*decision tree*). Input utama untuk algoritma C4.5 adalah dataset yang terdiri dari contoh-contoh pelatihan (*training examples*) yang berisi sekumpulan atribut dan label kelas yang sudah diketahui. Tujuan dari algoritma C4.5 adalah untuk membangun sebuah pohon keputusan yang dapat digunakan untuk mengklasifikasikan data baru berdasarkan atribut-atribut yang ada.

Pada langkah-langkah algoritma C4.5, algoritma akan memilih atribut yang paling informatif untuk digunakan sebagai node pemisah dalam membangun pohon keputusan.

Algoritma menggunakan metode pengukuran keuntungan informasi (*information gain*) untuk menentukan atribut yang paling baik. Selanjutnya, algoritma akan membagi dataset berdasarkan nilai-nilai atribut yang mungkin, dan langkah ini diulangi secara rekursif untuk setiap cabang yang dihasilkan hingga mencapai kondisi terminasi. Salah satu keunggulan utama dari algoritma C4.5 adalah kemampuannya untuk menangani atribut dengan nilai yang kontinu dan nilai yang hilang (*missing values*).

Input yang disebut sebagai "*samples*" merujuk pada data yang akan diklasifikasikan menggunakan pohon keputusan yang sudah dibangun. Setelah membangun pohon keputusan dengan algoritma C4.5, dapat menggunakan pohon tersebut untuk

mengklasifikasikan data baru berdasarkan atribut-atribut yang ada.

Rumus perhitungan nilai Gain adalah:

$$Gain(S,A) = Entropy(S) - \sum_{i=0}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

S = Himpunan Kasus.

A = Atribut.

I = Jumlah Partisi Atribut.

|S_i| = Jumlah Kasus pada partisi ke i.

|S| = Jumlah Kasus dalam S.

Rumus menghitung nilai Entropy adalah:

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

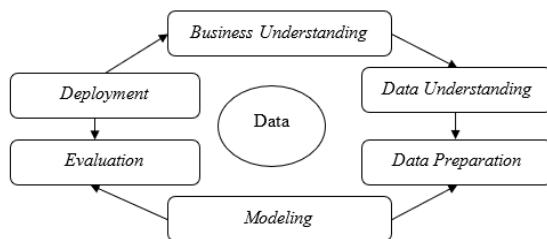
S = Himpunan Kasus

n = Jumlah partisi S

p_i = Proporsi dari S_i terhadap S

3. Metode Analisis Data

Penelitian ini menggunakan Cross Industry Standard Process for Data Mining (CRISP-DM) sebagai pendekatan sistematis dalam mengelola data mining. Tahap-tahap dalam model CRISP-DM yang digunakan dalam penelitian ini adalah sebagai berikut:



Sumber: Hasil Penelitian (2023)
Gambar 1. Model CRISP-DM

- Pemahaman Bisnis:**
Mengetahui tingkat kelulusan peserta didik kelas IX di SMP Negeri 184 Jakarta dengan algoritma C4.5.
- Pemahaman Data:**
Kumpulkan data dari SMP Negeri 184 Jakarta untuk menentukan atribut dan membuat dataset yang diperlukan.
- Persiapan Data:**
Perbaiki dan bersihkan data, termasuk memilih variabel penting, menghapus nilai yang hilang, dan menggabungkan data dari berbagai sumber.
- Pemodelan:**
Terapkan algoritma C4.5 untuk membangun model prediksi menggunakan teknik cross-validation pada data yang telah dipersiapkan.
- Evaluasi:**
Evaluasi model menggunakan *Confusion Matrix* untuk mengukur akurasi dan kinerja, mendapatkan informasi prediksi yang benar dan salah.

f. Penyajian:

Sajikan hasil analisis dan temuan dari algoritma C4.5 sebagai hasil penelitian.

Algoritma C4.5 diuji menggunakan *Cross Validation* pada aplikasi RapidMiner, sebuah platform analisis data yang dapat memperoleh estimasi kinerja yang lebih objektif dan dapat diandalkan dari model yang dibangun dengan algoritma C4.5. *Cross Validation* merupakan cara untuk memeriksa seberapa baik model bekerja. Ini dilakukan dengan membagi dataset menjadi beberapa bagian yang tumpang tindih, lalu menggunakan setiap bagian secara bergantian sebagai data uji.

Hasil dari pengujian menggunakan *Cross Validation* memperlihatkan kinerja algoritma C4.5 dengan lebih baik, dengan mencakup metrik-metrik evaluasi seperti akurasi, presisi, recall, dan area di bawah kurva ROC (AUC-ROC). Estimasi kinerja ini membantu memberikan gambaran yang lebih mendalam tentang sejauh mana model pohon keputusan yang dihasilkan oleh algoritma C4.5 dapat memprediksi data baru dengan tingkat keandalan yang tinggi.

III. HASIL DAN PEMBAHASAN

1. Dataset

Dataset yang digunakan berupa nilai peserta didik kelas IX tahun 2022-2023 SMP Negeri 184 Jakarta sebanyak 355 peserta didik yang sebelumnya sudah diolah dengan mempersempit ruang lingkup penilaian kelulusan dengan mengambil nilai rata-rata rapor dari semester 1 sampai semester 6, dan mengambil nilai rata-rata ujian sekolah dari 10 mata pelajaran, serta jumlah mata pelajaran yang remedial (her) yang diambil dari data mentah nilai ijazah dengan ketentuan Skor Ketuntasan Minimal (SKM) 81.50. Sementara untuk nilai karakter merupakan nilai olahan dan penilaian kelulusan yang akan diolah (dataset) secara keseluruhan menggunakan Skor Ketuntasan Minimal (SKM) 83.50.

No	Nama	Nilai Karakter	Nilai Ujian Sekolah	Nilai Rapor	Jumlah Her	Keterangan
1	Ahmad Hafiz Febriyanto	Baik	85.02	87.22	0	LULUS
2	Alexander Zakaria Sinanjuntak	Cukup	80.69	83.75	3	TIDAK LULUS
3	Alfy Nikardi	Kurang	82.52	84.55	1	TIDAK LULUS
4	Alysha Hikmaya Rachman	Baik	81.78	84.83	0	TIDAK LULUS
5	Ari Maulana	Cukup	80.13	83.30	>3	TIDAK LULUS
6	Arthur Maher Guterres	Baik	85.23	86.18	1	LULUS
7	Dendy Logosta	Baik	84.93	88.01	1	LULUS
8	Dhavina Nazwa Adliyani	Baik	89.76	91.32	0	LULUS
9	Dhava Dwi Rizki	Baik	85.10	86.22	0	LULUS
10	Early Pusпита Hermawan	Baik	85.49	87.30	1	LULUS
11	Fathia Litaqumalkarimah	Baik	85.94	86.16	0	LULUS
12	Gabriela Feandra	Baik	87.48	88.77	0	LULUS
13	Gahub Ganiya Aqila	Baik	83.63	85.95	0	LULUS
14	Grifania Anabela Gultom	Baik	89.08	90.09	0	LULUS
15	Gloria Ratuwella	Baik	89.29	90.47	0	LULUS
16	Ibnu Rasya	Kurang	78.25	84.32	>3	TIDAK LULUS
17	Javid Rabbani	Cukup	82.50	84.47	0	TIDAK LULUS
18	Jingga Laila Ismawardhani	Cukup	82.53	85.04	2	TIDAK LULUS
19	Keeyra Amanda Putri	Baik	84.24	87.11	0	LULUS

Sumber: Hasil Penelitian (2023)
Gambar 2. Sampel data

2. Algoritma C4.5 (decision tree)

Dalam membangun *decision tree* atau pohon keputusan, perlu dilakukan tahapan-tahapan dengan proses perhitungan. Tahap awal adalah menentukan jumlah class bagi peserta didik yang lulus atau tidak lulus. Selanjutnya, menghitung nilai entropy total, nilai entropy dari setiap atribut, menghitung nilai gain ratio, dan menentukan gain tertinggi. Dari data training yang ada, dengan total keseluruhan sebanyak 355 siswa, diketahui jumlah siswa yang lulus adalah 232 siswa, dan yang tidak lulus adalah 123 siswa. Sehingga diperoleh nilai *entropy* total:

$$\begin{aligned} \text{Entropy (total)} &= \\ & (232/355 * \log_2(232/355)) + \\ & (123/355 * \log_2(123/355)) \\ \text{Entropy (total)} &= 0.930884017 \end{aligned}$$

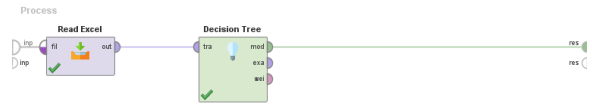
Proses untuk memperoleh gain dari setiap atribut melibatkan perhitungan entropi atribut berdasarkan kasus-kasus yang ada. Sehingga diperoleh hasil perhitungan entropy dan gain, dan menunjukkan atribut dengan nilai paling tinggi yaitu Nilai Karakter sebesar 0.621373801. Nilai Karakter merupakan atribut yang akan menjadi akar dari pohon keputusan (*decision tree*) dengan tiga cabang, yaitu Kurang, Cukup Baik.

Tabel III. 1
Hasil Perhitungan Entropy dan Gain

Atribut	Nilai Atribut	Jumlah (S)	LULUS	TIDAK LULUS	Entropy	Gain
Total		355	232	123	0,930884017	
Nilai Karakter	Kurang	42	0	42	0	0,621373801
	Cukup	56	4	52	0,371232327	
	Baik	257	228	29	0,508424658	
Nilai Ujian Sekolah	<83.50	134	23	111	0,661446756	0,552433558
	83.50 s.d 85	68	56	12	0,672294817	
	>85	153	153	0	0	
Nilai Rapor	<83.50	11	0	11	0	0,289864307
	83.50 s.d 85	56	3	53	0,301378644	
	>85	288	229	59	0,73154442	
Jumlah Her	0	225	205	20	0,432750159	0,470955386
	1	68	27	41	0,969204095	
	2	27	0	27	0	
	3	19	0	19	0	
	>3	16	0	16	0	

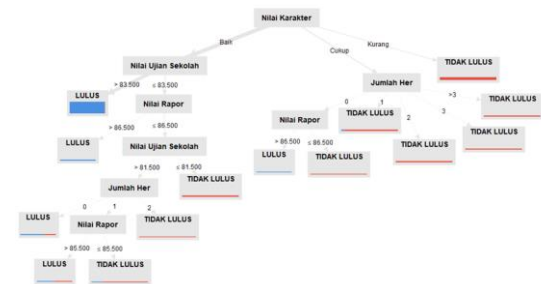
Sumber: Hasil Penelitian (2023)

Setelah perhitungan nilai entropy dan gain, pohon keputusan akan terbentuk dengan menggunakan aplikasi RapidMiner. Penerapan algoritma C4.5 yang akan membentuk pohon keputusan dengan RapidMiner dilakukan dengan menginput dataset berupa data nilai dalam bentuk excel, selanjutnya menerapkan model prediksi yaitu *decision tree*, dan menghubungkan dataset dengan *decision tree*.



Sumber: Hasil Penelitian, 2023
Gambar III. 2 Penerapan *Decision Tree* dalam RapidMiner

Sehingga pohon keputusan terbentuk.



Sumber: Hasil Penelitian (2023)
Gambar 3. Pohon keputusan menggunakan RapidMiner

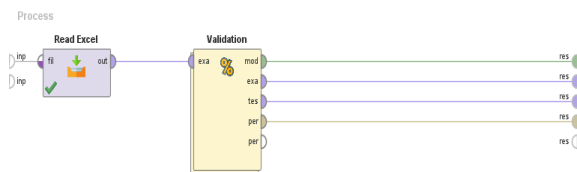
Dari pohon keputusan tersebut, algoritma C4.5 melakukan proses pengembangan rules atau aturan-aturan yang digunakan untuk melakukan klasifikasi data. Setiap jalur dari akar hingga daun pada pohon mewakili suatu *rule*. *Rules* atau aturan ini memberikan panduan dalam menentukan kategori atau kelas yang sesuai untuk suatu data baru berdasarkan nilai atribut-atribut yang dimiliki. Berikut adalah *rules* atau aturan yang dihasilkan pohon keputusan.

- Nilai Karakter = Baik
- Nilai Ujian Sekolah > 83.500: LULUS {LULUS=207, TIDAK LULUS=2}
- Nilai Ujian Sekolah ≤ 83.500
- Nilai Rapor > 86.500: LULUS {LULUS=11, TIDAK LULUS=0}
- Nilai Rapor ≤ 86.500
- Nilai Ujian Sekolah > 81.500
- Jumlah Her = 0: LULUS {LULUS=6, TIDAK LULUS=3}
- Jumlah Her = 1
- Nilai Rapor > 85.500: LULUS {LULUS=3, TIDAK LULUS=3}
- Nilai Rapor ≤ 85.500: TIDAK LULUS {LULUS=1, TIDAK LULUS=4}
- Jumlah Her = 2: TIDAK LULUS {LULUS=0, TIDAK LULUS=2}
- Nilai Ujian Sekolah ≤ 81.500: TIDAK LULUS {LULUS=0, TIDAK LULUS=15}
- Nilai Karakter = Cukup
- Jumlah Her = 0
- Nilai Rapor > 86.500: LULUS {LULUS=3, TIDAK LULUS=0}
- Nilai Rapor ≤ 86.500: TIDAK LULUS {LULUS=0, TIDAK LULUS=3}
- Jumlah Her = 1: TIDAK LULUS {LULUS=1,

TIDAK LULUS=15}
 Jumlah Her = 2: TIDAK LULUS {LULUS=0, TIDAK LULUS=15}
 Jumlah Her = 3: TIDAK LULUS {LULUS=0, TIDAK LULUS=9}
 Jumlah Her = >3: TIDAK LULUS {LULUS=0, TIDAK LULUS=10}
 Nilai Karakter = Kurang: TIDAK LULUS {LULUS=0, TIDAK LULUS=42}

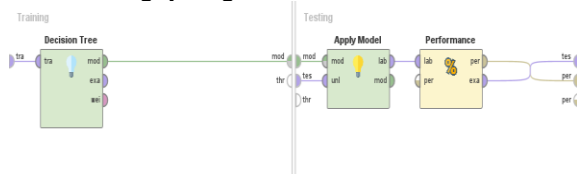
3. Pengujian Algoritma C4.5

Pengujian algoritma C4.5 dengan menggunakan *Cross Validation* dapat memperoleh estimasi kinerja yang lebih objektif dan dapat diandalkan dari model yang dibangun dengan algoritma C4.5.



Sumber: Hasil Penelitian (2023)
 Gambar 4. Penerapan Cross Validation

Dataset diuji menggunakan *Cross Validation* dan menghasilkan *data training* dengan model *Decision Tree*, selanjutnya *data training* akan ditesting dengan model *Apply Model* dan *%Performance (Classification)* pada aplikasi RapidMiner untuk menguji tingkat akurasi.



Sumber: Hasil Penelitian (2023)
 Gambar5. Proses dalam Cross Validation

4. Evaluasi Pengujian Algoritma C4.5

Evaluasi terhadap pengujian algoritma C4.5 menggunakan metode *Confusion Matrix* untuk mendapatkan nilai akurasi dan Kurva ROC (*Receiver Operating Characteristic*) untuk mengukur kinerja algoritma C4.5.

Confusion Matrix adalah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi. Tabel ini membandingkan nilai prediksi dengan nilai sebenarnya dari data yang diuji.

Tabel III. 1 Evaluasi Kinerja Klasifikasi

	True Lulus	True Tidak Lulus	Class precision
Pred. Lulus	224	15	93.72%
Pred. Tidak Lulus	8	108	93.10%
Class recall	96.55%	87.80%	

Sumber: Hasil Penelitian (2023)

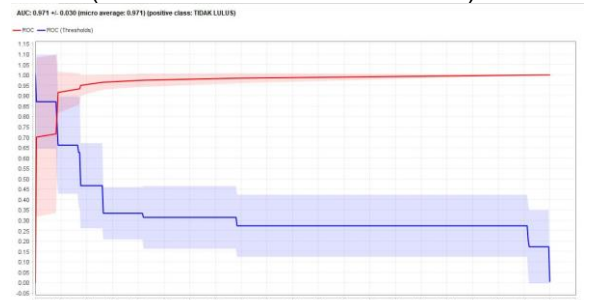
Dengan empat komponen utama, yaitu:

- True Positive (TP)*: Jumlah observasi yang benar diklasifikasikan sebagai positif. Dalam hal ini, TP berjumlah 224.
- True Negative (TN)*: Jumlah observasi yang benar diklasifikasikan sebagai negatif. Dalam hal ini, TN berjumlah 108.
- False Positive (FP)*: Jumlah observasi yang salah diklasifikasikan sebagai positif. Dalam hal ini, FP berjumlah 15.
- False Negative (FN)*: Jumlah observasi yang salah diklasifikasikan sebagai negatif. Dalam hal ini, FN berjumlah 8.

Pada Tabel III.1 dijelaskan bahwa kategori lulus memperoleh *class precision* 93.72%, dan *recall* atau *sensitivity* 96.55%, kategori tidak lulus memperoleh *class precision* 93.10% dan *class recall* 87.80%. Sementara itu, nilai *accuracy* dari perhitungan algoritma C4.5 adalah sebesar 93.54%.

Selanjutnya kurva ROC yang merupakan grafik yang menggambarkan kinerja algoritma klasifikasi pada berbagai threshold atau ambang batas. Pada kurva ROC, sumbu horizontal menggambarkan *False Positive Rate (FPR)*, yaitu rasio *false positive* terhadap total negatif, sedangkan sumbu vertikal menggambarkan *True Positive Rate (TPR)*, yaitu rasio *true positive* terhadap total positif.

Kurva ROC memberikan gambaran tentang kemampuan algoritma dalam membedakan antara kelas positif dan negatif. Nilai yang umum digunakan untuk mengukur performa dari kurva ROC adalah *Area Under the Curve (AUC)*. Nilai AUC yang mendekati 1 menunjukkan kinerja yang baik, sedangkan nilai mendekati 0.5 menunjukkan kinerja yang buruk (tidak lebih baik dari kebetulan).



Sumber: Hasil Penelitian (2023)
 Gambar 6. Kurva ROC

Dari kurva ROC dengan nilai AUC 0,971 menunjukkan bahwa kinerja algoritma klasifikasi dinilai sangat baik karena nilai AUC yang mendekati 1.

IV. KESIMPULAN

1. Tingkat kelulusan peserta didik kelas IX pada SMP Negeri 184 Jakarta dapat diklasifikasi dengan menerapkan teknik data mining dengan metode algoritma C4.5.
2. Setelah dilakukan pengujian menggunakan metode Cross Validation dengan aplikasi RapidMiner, diperoleh nilai akurasi sebesar 93.54% dari data keseluruhan sebanyak 355. Kategori lulus memperoleh class precision 93.72%, dan recall atau sensitivity 96.55%, kategori tidak lulus memperoleh class precision 93.10% dan class recall 87.80%. Hasil tersebut diperoleh dari perhitungan nilai true positive sebanyak 224 record, true negative sebanyak 108 record, false positive sebanyak 15 record, dan false negative sebanyak 8 record. Sementara itu, dalam kurva ROC, diperoleh nilai AUC sebesar 0,971 yang menunjukkan kinerja algoritma klasifikasi dinilai sangat baik.
3. Dengan adanya prediksi tingkat kelulusan yang akurat, diharapkan dapat memberikan informasi yang berharga bagi pihak sekolah dalam mengambil keputusan dan memberikan bimbingan kepada peserta didik untuk mencapai kelulusan yang sukses.

V. REFERENSI

Utomo, D. P., & Purba, B. (2019). Penerapan Data Mining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia.

- Prosiding Seminar Nasional Riset Information Science (SENARIS), 846–853.
- Azhar, Y., Firdausy, A. K., & Amelia, P. J. (2022). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke. *SINTECH Journal*, 5. <https://doi.org/10.31598>
- Etriyanti, E., Syamsuar, D., & Kunang, Y. N. (2020). Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa. *Telematika*, 13(1), 56–67. <https://doi.org/10.35671/telematika.v13i1.881>
- Muhamad, Windarto, A. P., & Suhada. (2019). PENERAPAN ALGORITMA C4.5 PADA KLASIFIKASI POTENSI SISWA DROP OUT. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 3(1), 753–760. <https://doi.org/10.30865/komik.v3i1.1688>
- Maulid, R. (2023, Januari 7). RapidMiner : Tools Data Science Andalan Data Experts. *DQLab*. <https://dqlab.id/rapidminer--tools-data-science-andalan-data-experts>
- Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression. *Jurnal Nasional Teknologi dan Sistem Informasi*, 7(1), 8–17. <https://doi.org/10.25077/teknosi.v7i1.2021.8-17>