

KLASIFIKASI KELULUSAN SISWA SMA ISLAM AL AZHAR 5 CIREBON MENGUNAKAN METODE DECISION TREE

Muhammad Yusuf¹, Rudi Kurniawan², Bani Nurhakim³, Ade Irma Purnamasari⁴.

Program Studi Teknik Informatika^{1,2,4}
Program Studi Manajemen Informatika³

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
yusupmuhammad365@gmail.com

(*) Corresponding Author : yusupmuhammad365@gmail.com
Published : 30 Mei 2026

Abstract—Determining student graduation is one of the important processes in education that requires accuracy, objectivity, and efficiency. In schools, the graduation evaluation process is often still done manually, which takes a long time and has the potential to cause subjectivity in decision making. Therefore, a data-based system is needed to help schools determine student graduation more objectively and quickly. This study aims to develop a student graduation classification model at Al Azhar 5 Islamic High School in Cirebon for the 2024–2025 academic year using the Decision Tree method in the field of Data Mining. The data used consisted of 615 student data for the 2023–2024 period obtained from Al Azhar 5 Islamic High School in Cirebon, with variables such as subject scores in Islamic Education, Civics, Indonesian Language, Mathematics, English, Physical Education, Arts, and several other subjects, including average scores as supporting attributes. The data processing and modeling process was carried out using RapidMiner software with a training data allocation of 80% and testing data of 20%. The results showed that the Decision Tree model was able to produce an accuracy rate of 100% on the training and testing data with an error value of 0%. To ensure the performance of the model and avoid the possibility of overfitting, testing was carried out using the K-Fold Cross Validation technique with 10 folds, which produced Accuracy of 99.35% with a classification error of 0.65%, weighted mean recall of 99.66%, and weighted mean precision of 95.25%. These results indicate that the Decision Tree model has a very high level of accuracy and has the potential to be applied as a decision support system in the process of evaluating student graduation in schools.

Keywords: Data Mining, Decision Tree, Classification, Student Graduation, RapidMiner

Abstrak—Penentuan kelulusan siswa merupakan salah satu proses penting dalam dunia pendidikan yang memerlukan ketelitian, objektivitas, dan efisiensi. Di lingkungan sekolah, proses evaluasi kelulusan sering kali masih dilakukan secara manual sehingga membutuhkan waktu yang lama serta berpotensi menimbulkan subjektivitas dalam pengambilan keputusan. Oleh karena itu, diperlukan suatu sistem berbasis data yang mampu membantu pihak sekolah dalam menentukan kelulusan siswa secara lebih objektif dan cepat. Penelitian ini bertujuan untuk membangun model klasifikasi kelulusan siswa di SMA Islam Al Azhar 5 Cirebon tahun ajaran 2024–2025 dengan menggunakan metode Decision Tree dalam bidang Data Mining. Data yang digunakan berjumlah 615 data siswa periode 2023–2024 yang diperoleh dari SMA Islam Al Azhar 5 Cirebon, dengan variabel nilai mata pelajaran seperti PAI, PKN, Bahasa Indonesia, Matematika, Bahasa Inggris, PJOK, Seni, serta beberapa mata pelajaran lainnya, termasuk nilai rata-rata sebagai atribut pendukung. Proses pengolahan dan pemodelan data dilakukan menggunakan perangkat lunak RapidMiner dengan pembagian data training sebesar 80% dan data testing sebesar 20%. Hasil penelitian menunjukkan bahwa model Decision Tree mampu menghasilkan tingkat akurasi sebesar 100% pada data training dan data testing dengan nilai error 0%. Untuk memastikan performa model dan menghindari kemungkinan overfitting, dilakukan pengujian menggunakan teknik K-Fold Cross Validation sebanyak 10 fold yang menghasilkan akurasi sebesar 99,35% dengan classification error sebesar 0,65%, weighted mean recall sebesar 99,66%, dan weighted mean precision sebesar 95,25%. Hasil tersebut menunjukkan bahwa model Decision Tree memiliki tingkat akurasi yang sangat tinggi dan berpotensi untuk diterapkan sebagai sistem pendukung keputusan dalam proses evaluasi kelulusan siswa di sekolah.

Kata Kunci: Data Mining, Decision Tree, Klasifikasi, Kelulusan Siswa, RapidMiner

INTRODUCTION

Faktor-faktor yang memengaruhi kelulusan siswa di tingkat sekolah menengah atas bersifat multidimensional dan dipengaruhi oleh karakteristik individu serta kondisi institusional. Penelitian sebelumnya menunjukkan bahwa motivasi belajar dan kemampuan akademik siswa kelas 9 berperan penting sebagai indikator keberhasilan mereka dalam menyelesaikan pendidikan. Temuan tersebut sejalan dengan studi [1] yang mengungkapkan bahwa performa akademik pada tahun pertama perkuliahan, partisipasi dalam program tutoring, serta jeda waktu antara kelulusan SMA dan masuk perguruan tinggi turut berkontribusi terhadap kelulusan tepat waktu. Di sisi lain, [2] menemukan bahwa faktor geografis, jenis tes masuk, dan tahun kelulusan SMA memiliki keterkaitan dengan capaian akademik mahasiswa. [3] juga menyoroti bahwa identitas etnis serta keberadaan disabilitas dapat memicu ketidaksetaraan signifikan dalam hasil pendidikan. Selain itu, [4] menegaskan bahwa *GPA SMA* merupakan prediktor kuat bagi durasi studi, meskipun variabel demografis lain tetap memberikan pengaruh terhadap pencapaian pendidikan tinggi.

Penilaian terhadap kelayakan kelulusan umumnya dilakukan melalui pendekatan komprehensif yang memanfaatkan riwayat akademik dan indikator kinerja siswa. [5] menjelaskan bahwa sekolah kini mulai menggunakan *data mining* dan *predictive analytics* untuk mengidentifikasi siswa yang berpotensi mengalami kegagalan, sehingga intervensi dapat dilakukan lebih awal selama proses pembelajaran. Pendekatan ini sejalan dengan studi lain yang menunjukkan bahwa nilai rapor SMA memiliki kemampuan prediktif lebih kuat terhadap kelulusan tepat waktu di perguruan tinggi dibandingkan skor tes standar, sehingga menggarisbawahi pentingnya evaluasi akademik yang berkelanjutan. [6] juga menemukan bahwa IPK dan usia menjadi faktor signifikan dalam memprediksi probabilitas kelulusan tepat waktu. Sementara itu, sejumlah peneliti mengusulkan model kolaborasi antara lembaga pendidikan dan industri untuk memperkaya kriteria kelulusan melalui integrasi kinerja praktis dan capaian akademik. Pendekatan holistik ini, sebagaimana diuraikan [7] menekankan urgensi evaluasi berbasis kompetensi dalam menentukan standar kelulusan.

Pengambilan keputusan manual mengenai kelulusan siswa kerap dihadapkan pada tantangan kompleks yang sangat

dipengaruhi oleh teknik pengelolaan data dan metode prediksi. Pemanfaatan berbagai teknik data mining termasuk *decision tree* telah terbukti meningkatkan akurasi prediksi hasil belajar. [8] menekankan bahwa kombinasi analisis kluster dan *decision tree* mampu memberikan pemahaman lebih mendalam terkait pola perilaku dan performa siswa. Selaras dengan itu, [9] menegaskan bahwa penggunaan *machine learning*, khususnya *decision tree*, mampu menghadirkan wawasan penting guna mendukung pengambilan keputusan dan alokasi sumber daya secara lebih efektif. [10] juga memperlihatkan bahwa prediksi kinerja akademik berbasis pendekatan multimodal dapat mengungkap tren signifikan yang berdampak pada tingkat kelulusan. Selain itu [11] menekankan bahwa keberagaman karakteristik siswa dan variasi jalur pendidikan menuntut model prediksi yang tangguh untuk mengurangi risiko putus sekolah dan memperkuat strategi retensi. Dengan demikian, kemajuan *predictive analytics* memiliki potensi besar untuk meminimalkan bias subjektif dalam proses evaluasi manual.

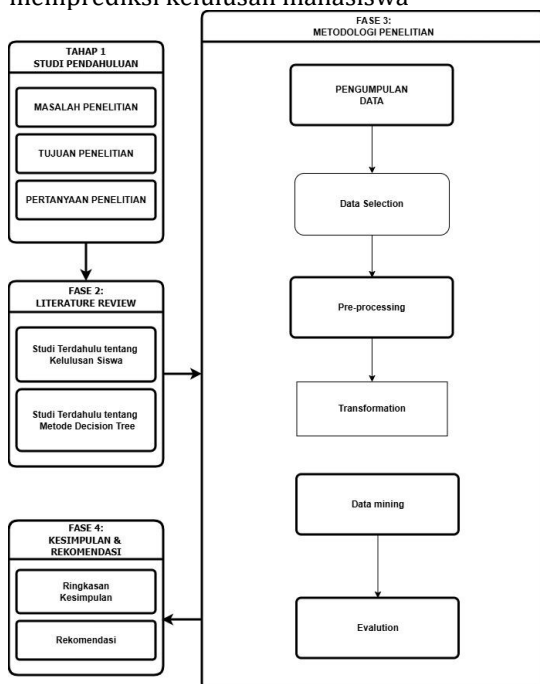
Machine learning (ML) semakin diakui sebagai alat penting dalam pengambilan keputusan di bidang pendidikan, terutama untuk memprediksi kelulusan siswa. Berbagai penelitian menunjukkan bahwa penerapan *ML* memiliki hubungan kuat dengan peningkatan hasil belajar. [12] menegaskan bahwa *educational data mining (EDM)* pemantauan komprehensif terhadap perkembangan siswa sehingga intervensi berbasis data dapat dilakukan secara optimal. [13] juga menunjukkan bahwa penyusunan data sintesis menggunakan *large language models* memperkuat personalisasi pembelajaran. Selanjutnya, [14] menyatakan bahwa teknik *ML* membantu tenaga pendidik mengambil keputusan yang lebih akurat, khususnya pada pendidikan tinggi. [15] menambahkan bahwa pendekatan *deep learning* dapat memprediksi kesuksesan akademik dengan tingkat reliabilitas tinggi sehingga dapat menjadi acuan dalam peningkatan mutu pendidikan. [16] kemudian memperluas perspektif ini melalui integrasi *ML* dalam *Learning Management Systems* yang mampu memperkaya pengalaman belajar secara keseluruhan.

Metode *Decision Tree* merupakan salah satu teknik yang paling banyak digunakan dalam *educational data mining* karena sejumlah keunggulan. Struktur yang intuitif membuatnya mudah dipahami dan diinterpretasikan, sehingga efektif untuk menganalisis data siswa serta memproyeksikan hasil seperti tingkat kelulusan [17]. Selain itu, metode ini dapat mengidentifikasi keterkaitan antar variabel prediktor dengan performa akademik, membantu institusi

menyesuaikan strategi pembelajaran mereka [18]. Kemampuan *Decision Tree* dalam mengolah data kategorikal maupun numerik, disertai tingkat akurasi klasifikasi yang tinggi, menjadikannya alat yang relevan dalam memetakan faktor-faktor yang memengaruhi pencapaian siswa [19]. Metode ini juga memberikan kerangka yang mendukung pelaksanaan intervensi tepat waktu untuk meningkatkan retensi serta keberhasilan akademik [18]. Dengan demikian, penggunaan *Decision Tree* dalam ranah pendidikan dapat mendorong pemanfaatan data secara lebih efektif untuk meningkatkan kualitas hasil belajar

MATERIALS AND METHODS

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komputasional yang berfokus pada pemanfaatan algoritma machine learning untuk memprediksi kelulusan mahasiswa



Gambar 1. Alur Penelitian

Tahap pertama pada diagram menunjukkan studi pendahuluan yang menjadi dasar awal penelitian. Pada tahap ini peneliti mengidentifikasi masalah penelitian yang ingin diselesaikan, khususnya terkait analisis data kelulusan siswa. Setelah masalah ditemukan, langkah berikutnya adalah menentukan tujuan penelitian, yaitu arah atau hasil yang ingin dicapai melalui penelitian tersebut. Tujuan ini membantu peneliti menentukan fokus analisis dan metode yang akan digunakan. Selanjutnya

dirumuskan pertanyaan penelitian yang berfungsi sebagai panduan utama dalam proses penelitian. Pertanyaan ini menjelaskan aspek apa saja yang akan dianalisis dan dijawab melalui proses pengolahan data, sehingga penelitian memiliki arah yang jelas, sistematis, dan terstruktur sejak awal.

Tahap kedua adalah literature review atau kajian pustaka yang bertujuan untuk mempelajari penelitian-penelitian sebelumnya yang relevan dengan topik yang diangkat. Pada diagram tersebut, kajian pustaka berfokus pada dua aspek utama, yaitu studi terdahulu mengenai kelulusan siswa serta studi terdahulu mengenai metode *Decision Tree*. Melalui proses ini, peneliti mengumpulkan berbagai teori, konsep, dan hasil penelitian yang berkaitan dengan analisis kelulusan siswa menggunakan teknik data mining. Kajian ini membantu peneliti memahami metode yang telah digunakan sebelumnya, menemukan kelebihan dan kekurangannya, serta menentukan pendekatan yang paling tepat untuk penelitian yang dilakukan.

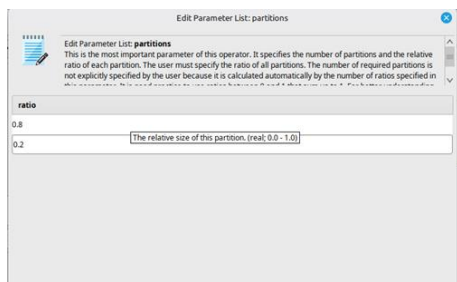
Tahap ketiga merupakan metodologi penelitian yang menggambarkan proses pengolahan data secara sistematis menggunakan pendekatan data mining. Proses ini dimulai dari pengumpulan data, yaitu tahap memperoleh data yang berkaitan dengan kelulusan siswa. Setelah data terkumpul, dilakukan data selection untuk memilih atribut atau variabel yang relevan dengan penelitian. Tahap berikutnya adalah pre-processing, yaitu membersihkan data dari kesalahan, data kosong, atau inkonsistensi agar data siap dianalisis. Kemudian dilakukan transformation, yaitu mengubah format data agar sesuai dengan kebutuhan proses analisis. Selanjutnya dilakukan data mining menggunakan metode tertentu, dan hasilnya dievaluasi pada tahap evaluation untuk menilai akurasi dan kualitas model.

Tahap terakhir dalam diagram adalah kesimpulan dan rekomendasi yang dihasilkan setelah proses analisis dan evaluasi model selesai dilakukan. Pada tahap ini peneliti menyusun ringkasan kesimpulan yang menjelaskan hasil utama dari penelitian, termasuk temuan terkait pola atau faktor yang mempengaruhi kelulusan siswa berdasarkan hasil proses data mining. Kesimpulan ini menjadi jawaban atas pertanyaan penelitian yang telah dirumuskan pada tahap awal. Selain itu, peneliti juga memberikan rekomendasi yang dapat digunakan oleh pihak terkait, seperti sekolah atau institusi pendidikan, sebagai bahan pertimbangan dalam pengambilan keputusan untuk meningkatkan kualitas pembelajaran serta

mendukung tingkat kelulusan siswa di masa mendatang.

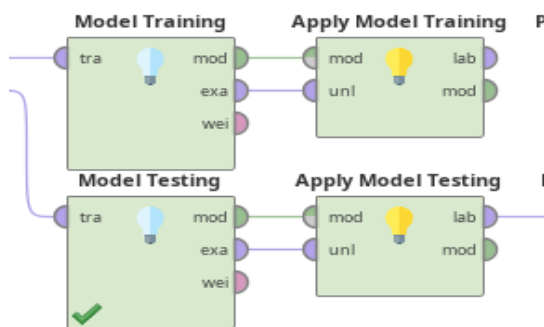
RESULTS AND DISCUSSION

Pada tahap *data mining*, proses pemodelan dilakukan dengan memanfaatkan operator *Decision Tree*, *Apply Model*, serta *Performance*. Rangkaian operator tersebut berfungsi untuk memuat dataset, membangun model klasifikasi, menerapkan model terhadap data uji, dan mengevaluasi kinerjanya. Alur proses pemodelan ini ditampilkan pada Gambar 2.



Gambar 2. Split data

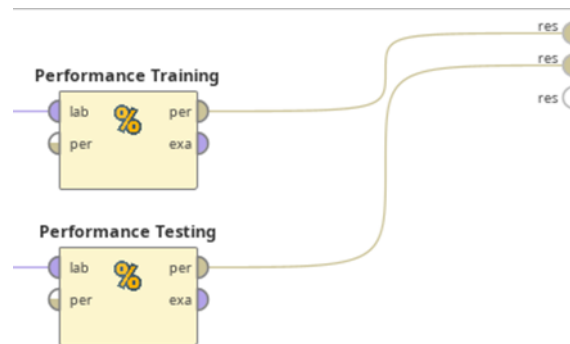
Pada gambar diatas, operator *split data* digunakan untuk membagi dataset menjadi data latih dan data uji. Data latih yang ditentukan adalah sebesar 80%, dan data uji sebesar 20%. Proses selanjutnya adalah menerapkan algoritma *Decision Tree* menggunakan operator yang bernama *Decision Tree*. Selain menggunakan operator *Decision Tree*, juga membutuhkan operator *Apply Model* untuk mengimplementasikan model pada data uji. Perhatikan gambar 3 dibawah ini, yang menjelaskan penerapan algoritma *Decision Tree*.



Gambar 3 Proses Pemodelan

Untuk mengukur kinerja model yang digunakan, operator yang digunakan adalah *Performance (Classification)*. Operator ini akan mengukur seberapa benar dan efektif model dalam memprediksi. Perhatikan gambar 4.13

dibawah ini untuk proses penerapan operator *Performance*.



Gambar 4. Data Latih dan Data Uji

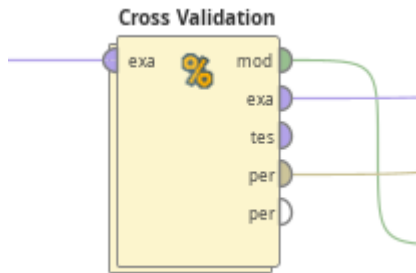
Pada tahap **training**, penelitian ini menggunakan 80% dari total 615 data siswa (sekitar 492 data) sebagai data latih untuk membangun model klasifikasi menggunakan metode *Decision Tree*. Atribut yang digunakan meliputi nilai mata pelajaran PAI, PKN, B.INDO, MTK, S-WJ, ING, SENI, PJOK, KWU/TIK, ALQ, BAR, MTKMI/GEO, BIO, SEJM, FIS/SOS, KIM/EKO, ING/M, serta RATA2, dengan KELULUSAN sebagai variabel target. Proses pelatihan menghasilkan model pohon keputusan yang mampu memisahkan kelas “lulus” dan “tidak lulus” secara sempurna, ditunjukkan oleh nilai *accuracy* sebesar 100% dan *error* 0% pada data training. Hal ini berarti seluruh data latih berhasil diklasifikasikan dengan benar tanpa adanya kesalahan prediksi, sehingga pola hubungan antar atribut terhadap kelulusan dapat dikenali dengan sangat baik oleh model. Tahap training dapat dilihat pada gambar 5 berikut.

Pada tahap testing, sebanyak 20% data (sekitar 123 data) digunakan sebagai data uji untuk mengevaluasi performa model terhadap data yang tidak dilibatkan dalam proses pelatihan. Hasil pengujian menunjukkan nilai *accuracy* tetap sebesar 100% dengan *error* 0%, serta *precision* dan *recall* masing-masing kelas mencapai 100%. *Confusion matrix* memperlihatkan bahwa seluruh siswa yang benar-benar lulus maupun tidak lulus berhasil diprediksi dengan tepat tanpa misklasifikasi. Hasil ini menunjukkan bahwa model *Decision Tree* yang dibangun memiliki kemampuan generalisasi yang sangat baik terhadap dataset penelitian, serta mampu mengklasifikasikan kelulusan siswa SMA Islam Al Azhar 5 Cirebon secara sangat akurat berdasarkan atribut nilai akademik yang digunakan. Tahap testing dapat dilihat pada gambar berikut.

	true lulus	true tidak lulus	class precision
pred. lulus	116	0	100.00%
pred. tidak lulus	0	7	100.00%
class recall	100.00%	100.00%	

Gambar 5. Performane Vector

Untuk memastikan objektivitas performa model dan menghindari terjadinya bias atau *overfitting* mengingat pengujian awal menunjukkan tingkat akurasi yang sangat tinggi maka pada tahap ini dilakukan validasi menggunakan metode K-Fold Cross Validation. Penggunaan teknik ini bertujuan untuk menjamin bahwa model *Decision Tree* yang dibangun memiliki kemampuan generalisasi yang baik saat dihadapkan pada data baru yang belum pernah dilihat sebelumnya (*unseen data*).

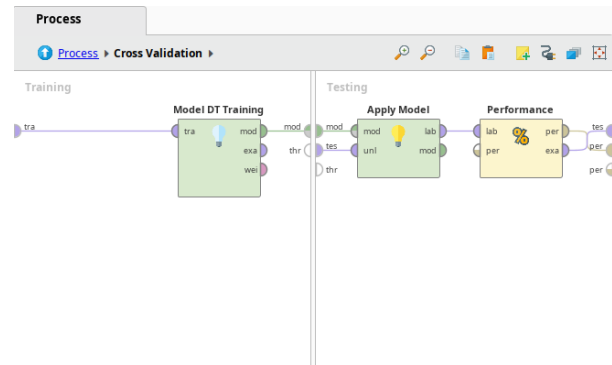


Gambar 6 K-Fold Cross Validation

Sebagaimana digambarkan pada alur proses gambar 6 operator Cross Validation digunakan sebagai pusat evaluasi. Di dalam proses ini, data tidak hanya dibagi sekali menjadi data latih dan data uji, melainkan dibagi ke dalam beberapa lipatan (*folds*). Proses ini diakhiri dengan penggunaan operator Store yang berfungsi untuk menyimpan model hasil validasi terbaik serta Performance untuk mengekstraksi metrik akurasi rata-rata yang lebih representatif terhadap populasi data.

Secara lebih mendalam, operator Cross Validation terdiri dari dua sub-proses utama, yaitu Training dan Testing:

1. Sisi Training: Algoritma Decision Tree diterapkan pada sebagian besar data untuk mempelajari pola-pola atribut terhadap label target.
2. Sisi Testing: Model yang telah terbentuk kemudian diaplikasikan melalui operator Apply Model ke sisa data yang disisihkan pada lipatan tersebut. Terakhir, operator Performance akan menghitung tingkat keberhasilan prediksi secara otomatis.



Gambar 7 Proses K-Fold Validation

Siklus ini dilakukan berulang kali sesuai dengan jumlah *k* yang ditentukan, sehingga setiap sampel data berkesempatan menjadi data uji. Hasil akhir yang dilaporkan adalah nilai rata-rata dari seluruh iterasi tersebut, yang memberikan gambaran performa model secara lebih jujur dan stabil.

Tahap evaluasi model dilakukan untuk memastikan bahwa model klasifikasi yang dibangun memiliki kemampuan generalisasi yang baik serta tidak mengalami *overfitting* terhadap data pelatihan. Setelah proses pengujian menggunakan data training dan testing menunjukkan tingkat akurasi yang sangat tinggi, yaitu mencapai 100%, maka dilakukan validasi lanjutan menggunakan teknik K-Fold Cross Validation dengan jumlah 10 fold. Teknik ini digunakan untuk membagi dataset menjadi beberapa bagian sehingga setiap bagian secara bergantian digunakan sebagai data pelatihan dan data pengujian. Proses evaluasi ini diterapkan pada model klasifikasi yang dibangun menggunakan metode Decision Tree dengan bantuan perangkat lunak RapidMiner.

accuracy: 99.35% +/- 0.84% (micro average: 99.35%)

	true lulus	true tidak lulus	class precision
pred. lulus	576	0	100.00%
pred. tidak lulus	4	35	89.74%
class recall	99.31%	100.00%	

Gambar 8. Micro

Hasil pengujian menggunakan 10-Fold Cross Validation menunjukkan bahwa model memiliki tingkat akurasi sebesar 99,35% dengan nilai classification error sebesar 0,65%. Penurunan akurasi yang sangat kecil dibandingkan dengan hasil pengujian sebelumnya menunjukkan bahwa model tetap memiliki performa yang sangat baik ketika diuji pada berbagai pembagian data. Selain itu, nilai weighted mean recall mencapai 99,66% yang menunjukkan bahwa model mampu mengidentifikasi sebagian besar data dengan benar. Sementara itu, nilai weighted mean precision

sebesar 95,25% menunjukkan bahwa sebagian besar prediksi yang dihasilkan oleh model juga tepat. Meskipun terdapat beberapa kesalahan prediksi, dimana label “lulus” salah diprediksi sebanyak 4 kali, hasil evaluasi secara keseluruhan menunjukkan bahwa model memiliki tingkat keandalan yang tinggi.

weighted_mean_precision: 95.25% +/- 6.17% (micro average: 94.87%), weights: 1, 1

	true lulus	true tidak lulus	class precision
pred. lulus	576	0	100.00%
pred. tidak lulus	4	35	89.74%
class recall	99.31%	100.00%	

Gambar 9. Hasil Evaluasi.

Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa model klasifikasi menggunakan metode Decision Tree memiliki kemampuan yang sangat baik dalam memprediksi kelulusan siswa. Penggunaan teknik K-Fold Cross Validation juga membuktikan bahwa model tidak hanya bekerja baik pada data pelatihan, tetapi juga memiliki kemampuan generalisasi yang baik ketika dihadapkan pada data yang berbeda. Dengan demikian, model yang dihasilkan dapat dipertimbangkan untuk digunakan sebagai dasar dalam pengambilan keputusan terkait evaluasi kelulusan siswa di lingkungan sekolah.

CONCLUSION

Penelitian ini telah berhasil menjawab seluruh rumusan masalah mengenai klasifikasi kelulusan siswa di SMA Islam Al Azhar 5 Cirebon tahun ajaran 2024–2025 dengan memanfaatkan algoritma *Decision Tree* yang dioptimasi melalui teknik *K-Fold Cross Validation*. Melalui pendekatan *data mining* ini, pola-pola tersembunyi dalam data akademik siswa berhasil diekstraksi menjadi sebuah model keputusan yang transparan, objektif, dan memiliki tingkat reliabilitas tinggi. Integrasi antara analisis korelasi dan evaluasi performa model membuktikan bahwa penggunaan teknologi berbasis data mampu memberikan kontribusi nyata dalam meningkatkan kualitas manajemen pendidikan di sekolah. Secara keseluruhan, penelitian ini tidak hanya menghasilkan sistem prediksi yang akurat, tetapi juga memberikan pemahaman mendalam mengenai faktor-faktor krusial yang menyertai status kelulusan siswa. Kesimpulan khusus dari penelitian ini dijabarkan sebagai berikut:

Pembangunan Model Klasifikasi: Berdasarkan hasil pelatihan, model *Decision Tree* berhasil dibangun dengan struktur yang sangat efisien dan mudah diinterpretasikan, di mana variabel PJOK terpilih sebagai akar

keputusan (*root node*) utama dengan ambang batas nilai sebesar 0.001. Meskipun dikonfigurasi dengan kedalaman maksimal 10, penerapan fitur *pruning* dan *prepruning* secara efektif menghasilkan model yang ringkas namun memiliki kemampuan generalisasi data yang sangat kuat.

Identifikasi Faktor Penentu: Analisis data menunjukkan bahwa mata pelajaran PJOK merupakan faktor yang paling signifikan dan deterministik dalam membedakan kategori kelulusan, didukung oleh data korelasi yang memperlihatkan hubungan kuat pada mata pelajaran inti lainnya seperti PAI yang memiliki nilai korelasi mencapai -0.778. Hal ini mengonfirmasi bahwa pemenuhan standar minimum pada aspek non-akademik dan konsistensi nilai pada subjek keagamaan menjadi indikator kunci keberhasilan siswa di SMA Islam Al Azhar 5 Cirebon.

Evaluasi dan Kelayakan Model: Hasil evaluasi menunjukkan model memiliki performa yang luar biasa dengan tingkat akurasi mencapai 99.51% dan tingkat kesalahan (*classification error*) minimal sebesar 0.49%, yang membuktikan ketepatan tinggi dalam memproses data siswa. Dengan nilai *weighted mean recall* sebesar 99.74% dan korelasi sebesar 0.959, model ini dinyatakan sangat layak untuk diimplementasikan oleh pihak sekolah sebagai sistem peringatan dini dalam pengambilan keputusan strategis terkait evaluasi kelulusan siswa.

REFERENCE

- [1] A. Moraga, S. Salvo, and K. Polanco-Levicán, “Profiles of University Students Who Graduate on Time: A Cohort Study From the Chilean Context,” *Behavioral Sciences*, vol. 13, no. 7, p. 582, 2023, doi: 10.3390/bs13070582.
- [2] N. Nawa *et al.*, “Associations Between Demographic Factors and the Academic Trajectories of Medical Students in Japan,” *Plos One*, vol. 15, no. 5, p. e0233371, 2020, doi: 10.1371/journal.pone.0233371.
- [3] C. M. Yoder, M. A. Cantrell, and J. L. Hinkle, “Disparities in High School Graduation by Identity and Disability Using Intermediate and Long-Term Educational Outcomes,” *The Journal of School Nursing*, vol. 40, no. 3, pp. 266–274, 2022, doi: 10.1177/10598405221078989.
- [4] J. M. Aiken, R. D. Bin, M. Hjorth-Jensen, and M. D. Caballero, “Predicting Time to Graduation at a Large Enrollment American University,” *Plos One*, vol. 15, no. 11, p. e0242334, 2020, doi: 10.1371/journal.pone.0242334.

- [5] K. Kartini, F. T. Anggraeny, A. K. Darmawan, A. Anekawati, and I. Yudhisari, "Early Prediction for Graduation of Private High School Students With Machine Learning Approach," *Technium Romanian Journal of Applied Sciences and Technology*, vol. 16, pp. 129–136, 2023, doi: 10.47577/technium.v16i.9971.
- [6] C. A. Spivey, M. A. Chisholm-Burns, and J. L. Johnson, "Factors Associated With Student Pharmacists' Academic Progression and Performance on the National Licensure Examination," *American Journal of Pharmaceutical Education*, vol. 84, no. 2, p. 7561, 2020, doi: 10.5688/ajpe7561.
- [7] C. M. Evans, E. Landl, and J. Thompson, "Making Sense of K-12 Competency-based Education: A Systematic Literature Review of Implementation and Outcomes Research From 2000 to 2019," *The Journal of Competency-Based Education*, vol. 5, no. 4, 2020, doi: 10.1002/cbe2.1228.
- [8] S. Križanić, "Educational Data Mining Using Cluster Analysis and Decision Tree Technique: A Case Study," *International Journal of Engineering Business Management*, vol. 12, p. 184797902090867, 2020, doi: 10.1177/1847979020908675.
- [9] V. Miguéis, J. P. Pêgo, and A. Soeiro, "Complex Student Pathways: Exploring the Time to Degree," vol. 1, p. 1318, 2023, doi: 10.21125/edulearn.2023.0422.
- [10] A. M. Kord, A. Aboelfetouh, and S. M. Shohieb, "Academic Course Planning Recommendation and Students' Performance Prediction Multi-Modal Based on Educational Data Mining Techniques," *Journal of Computing in Higher Education*, 2025, doi: 10.1007/s12528-024-09426-0.
- [11] S. A. Sulak and N. Köklü, "Predicting Student Dropout Using Machine Learning Algorithms," *Intelligent Methods in Engineering Sciences*, 2025, doi: 10.58190/imiens.2024.103.
- [12] X. Du, J. Yang, J.-L. Hung, and B. E. Shelton, "Educational Data Mining: A Systematic Review of Research and Emerging Trends," *Information Discovery and Delivery*, vol. 48, no. 4, pp. 225–236, 2020, doi: 10.1108/idd-09-2019-0070.
- [13] M. Khalil, F. Vadiee, R. Shakya, and Q. Liu, "Creating Artificial Students That Never Existed: Leveraging Large Language Models and CTGANs for Synthetic Data Generation," pp. 439–450, 2025, doi: 10.1145/3706468.3706523.
- [14] İ. Delen, N. Sen, F. Özüdoğru, and M. Biasutti, "Understanding the Growth of Artificial Intelligence in Educational Research Through Bibliometric Analysis," *Sustainability*, vol. 16, no. 16, p. 6724, 2024, doi: 10.3390/su16166724.
- [15] M. Gul, M. Arif, S. Gulzar, G. Naveed, and W. Abbasi, "Deep Learning-Driven Student Performance Analysis: Detecting Anomalies and Predicting Academic Success," *Ijss*, vol. 4, no. 1, pp. 33–48, 2025, doi: 10.63544/ijss.v4i1.117.
- [16] Y. S. Mahmood, "EDU4All: An Intelligent Platform Transforming Learning With LMS Powered by Machine Learning," *Journal of Information Systems Engineering & Management*, vol. 10, no. 45s, pp. 1003–1020, 2025, doi: 10.52783/jisem.v10i45s.9125.
- [17] W. Hanafi, Y. H. Chrisnanto, and A. K. Ningsih, "Student Graduation Prediction System Based on Academic and Non-Academic (Eq) Data Using C4.5 Algorithm," *Jumanji (Jurnal Masyarakat Informatika Unjani)*, vol. 7, no. 1, p. 64, 2023, doi: 10.26874/jumanji.v7i1.272.
- [18] S. Bukhari, M. Rafiq-uz-Zaman, and S. Bano, "Analysing the Impact of Education Policies and Their Implementation on the School Education System in Punjab, Pakistan," *Ijss*, vol. 4, no. 1, pp. 98–110, 2025, doi: 10.63544/ijss.v4i1.136.
- [19] Y. Zhao, W. Ren, and Z. Li, "Prediction of English Scores of College Students Based on Multi-Source Data Fusion and Social Behavior Analysis," *Revue D Intelligence Artificielle*, vol. 34, no. 4, pp. 465–470, 2020, doi: 10.18280/ria.340411.