

Data Mining untuk Memprediksi Kelulusan Mahasiswa Jurusan DIII Kebidanan Menggunakan Metode Klasifikasi C4.5 (Studi Kasus : Institut Teknologi dan Kesehatan Ika Bina)

Intan Nur Fitriyani¹, Sentosa Pohan², Desi Irpan³, Riswan Syahputra Damanik⁴, Baginda Restu Al Ghazali⁵

^{1,4,5}Sistem Informasi, Institut Teknologi dan Kesehatan Ika Bina, Rantauprapat, Indonesia

^{2,3}Teknologi Informasi, Fakultas Ilmu Komputer, Institut Teknologi dan Kesehatan Ika Bina

Email: ¹intannurfitriyani732@gmail.com, ²sentosa.pohan88@itkes-ikabina.ac.id, ³desi.irfan@itkes-ikabina.ac.id, ⁴syahp2304@gmail.com, ⁵bagindarestu123@gmail.com

Abstrak– Penelitian ini bertujuan untuk membangun model prediksi kelulusan mahasiswa pada program studi DIII Kebidanan di Institut Teknologi dan Kesehatan Ika Bina menggunakan algoritma C4.5. Metode yang digunakan adalah data mining dengan tahapan pengumpulan data, pembersihan data, penanganan nilai hilang, transformasi data, pemodelan menggunakan algoritma C4.5, dan evaluasi model. Data penelitian terdiri dari 150 mahasiswa, mencakup nilai mata kuliah, kehadiran, keterampilan klinis, kedisiplinan, pengalaman praktik, dan IPK. Hasil penelitian menunjukkan bahwa model prediksi yang dihasilkan mampu mengklasifikasikan kelulusan mahasiswa dengan akurasi 98%, precision 97%, recall 99%, dan AUC 0.97. Faktor utama yang berpengaruh terhadap kelulusan mahasiswa adalah nilai mata kuliah, kehadiran, keterampilan klinis, dan pengalaman praktik. Penelitian ini menegaskan bahwa penerapan data mining dengan algoritma C4.5 dapat digunakan sebagai alat bantu dalam proses evaluasi akademik dan pengambilan keputusan di perguruan tinggi vokasi, khususnya dalam program studi Kebidanan. Model prediksi yang dihasilkan juga dapat digunakan sebagai dasar untuk merancang intervensi yang lebih tepat bagi mahasiswa yang berisiko tidak lulus.

Kata Kunci: Kelulusan Mahasiswa, Data Mining, C4.5, Decision Tree, Kebidanan, Evaluasi Akademik

Abstract– This research aims to develop a predictive model of student graduation in the DIII Midwifery Study Program at Institut Teknologi dan Kesehatan Ika Bina using the C4.5 algorithm. The research method applies data mining techniques with stages including data collection, data cleaning, handling missing values, data transformation, modeling using the C4.5 algorithm, and model evaluation. The dataset consists of 150 students, covering variables such as course grades, attendance, clinical skills, discipline, practical experience, and GPA. The results indicate that the developed prediction model can classify student graduation status with an accuracy of 98%, precision of 97%, recall of 99%, and AUC of 0.97. The main factors affecting student graduation are course grades, attendance, clinical skills, and practical experience. This study confirms that the application of data mining using the C4.5 algorithm can be used as a supporting tool in the academic evaluation process and decision-making at vocational higher education institutions, particularly in midwifery study programs. The resulting prediction model can also serve as a basis for designing more targeted interventions for students at risk of not graduating.

Keywords: Student Graduation, Data Mining, C4.5, Decision Tree, Midwifery, Academic Evaluation

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah memberikan dampak signifikan di berbagai bidang, termasuk dalam dunia pendidikan. Salah satu aspek yang semakin mendapatkan perhatian adalah pemanfaatan data untuk mendukung pengambilan keputusan. Dalam konteks pendidikan tinggi, khususnya di perguruan tinggi vokasi, pemanfaatan data untuk memprediksi kelulusan mahasiswa merupakan salah satu aplikasi yang dapat mengoptimalkan proses evaluasi akademik. *Data mining*, yang merupakan teknik untuk mengekstraksi informasi yang berguna dari data besar, menjadi salah satu metode yang banyak digunakan untuk memprediksi kelulusan mahasiswa [1]. Salah satu metode dalam *data mining* yang sering diterapkan untuk klasifikasi adalah C4.5, sebuah algoritma pohon keputusan yang telah terbukti efektif dalam berbagai aplikasi klasifikasi [2], [3], [4].

Penerapan *data mining* dalam pendidikan telah banyak diteliti oleh berbagai peneliti sebelumnya. Salah satu penelitian yang relevan adalah yang dilakukan oleh Riyanto *et al.* [2] yang menggunakan algoritma C4.5 untuk memprediksi kelulusan mahasiswa pada program studi tertentu di Indonesia. Penelitian tersebut menunjukkan bahwa algoritma C4.5 mampu memberikan akurasi yang tinggi dalam memprediksi kelulusan mahasiswa dengan menggunakan variabel seperti nilai mata kuliah, kehadiran, dan kegiatan ekstrakurikuler. Selain itu, Nurahman *et al.* [3] juga menggunakan algoritma C4.5 dalam memprediksi kelulusan mahasiswa di bidang teknik, dengan fokus pada faktor-faktor akademik seperti IPK, kehadiran, dan partisipasi dalam kegiatan pembelajaran. Penelitian serupa juga dikemukakan oleh Sugianto dan Maulana [4], serta Yusuf dan Yuliana [5] yang membandingkan kinerja beberapa algoritma *machine learning*.

Kurniawan dan Haryanto [6], serta Sari dan Setiawan [7] menekankan pentingnya penerapan algoritma pohon keputusan dalam konteks pendidikan vokasi. Wahyu dan Widodo [8] memberikan tinjauan umum mengenai efektivitas berbagai metode *data mining* dalam memprediksi keberhasilan akademik. Kajian oleh Kurniawati dan Rizki [9] juga menunjukkan relevansi algoritma *machine learning* untuk memprediksi performa akademik.

Namun, meskipun banyak penelitian yang telah dilakukan, sebagian besar dari mereka masih terbatas pada jurusan atau program studi tertentu, terutama di perguruan tinggi dengan jurusan teknik atau sains. Penelitian mengenai penggunaan *data mining* untuk memprediksi kelulusan mahasiswa di program studi non-teknik, seperti DIII Kebidanan, masih sangat terbatas. Program studi Kebidanan memiliki karakteristik tersendiri, yang mencakup pengajaran teori dan praktik yang berorientasi pada profesi [10]. Oleh karena itu, penting untuk melakukan penelitian yang lebih mendalam mengenai penggunaan *data mining* dalam memprediksi kelulusan mahasiswa Kebidanan, dengan mempertimbangkan faktor-faktor yang relevan seperti keterampilan klinis, kedisiplinan, dan interaksi mahasiswa dengan dosen atau instruktur praktik. Program studi Kebidanan tidak hanya memerlukan kemampuan akademik tetapi juga keterampilan praktis yang sangat penting untuk kelulusan mereka [11], sehingga penggunaan *data mining* untuk memprediksi kelulusan menjadi sangat relevan untuk meningkatkan kualitas pendidikan.

Penelitian sebelumnya lebih banyak berfokus pada faktor akademik murni seperti nilai dan kehadiran mahasiswa [2], [3], [6]. Akan tetapi, untuk program studi Kebidanan, faktor-faktor non-akademik seperti keterampilan klinis, kedisiplinan, dan pengalaman praktik sangat berpengaruh terhadap kelulusan mahasiswa. Keterampilan klinis, yang mencakup kemampuan mahasiswa dalam praktik di lapangan, memegang peranan penting dalam menentukan kelulusan mereka dalam program studi Kebidanan. Kedisiplinan mahasiswa dalam mengikuti jadwal kuliah, praktikum, dan ujian juga menjadi faktor penting yang menentukan keberhasilan mereka. Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengembangkan model prediksi kelulusan mahasiswa program studi DIII Kebidanan dengan menggunakan algoritma *C4.5* dan mempertimbangkan faktor akademik serta non-akademik yang berpengaruh [12].

Penelitian ini bertujuan untuk memberikan kontribusi dalam pengembangan metode prediksi kelulusan yang lebih akurat dan komprehensif, dengan menggunakan algoritma *C4.5* untuk menganalisis data mahasiswa program studi DIII Kebidanan di Institut Teknologi dan Kesehatan Ika Bina. Penelitian ini juga bertujuan untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap kelulusan mahasiswa di program studi Kebidanan. Beberapa faktor yang akan dianalisis meliputi nilai mata kuliah, kehadiran, keterampilan klinis, dan kedisiplinan mahasiswa dalam mengikuti praktikum dan ujian. Dengan demikian, penelitian ini diharapkan dapat menghasilkan model prediksi yang dapat memberikan gambaran yang lebih jelas mengenai faktor-faktor yang mempengaruhi kelulusan mahasiswa serta membantu pihak perguruan tinggi dalam melakukan evaluasi dan intervensi terhadap mahasiswa yang membutuhkan dukungan tambahan [13].

Selain itu, penelitian ini juga bertujuan untuk mengevaluasi akurasi model yang dihasilkan dengan menggunakan algoritma *C4.5*. Akurasi model ini akan diuji menggunakan teknik *cross-validation* [14] untuk memastikan bahwa model prediksi yang dihasilkan dapat digunakan dengan baik dalam memprediksi kelulusan mahasiswa di masa depan. Dengan menggunakan metode ini, diharapkan hasil penelitian ini dapat memberikan wawasan baru tentang bagaimana *data mining* dapat diterapkan dalam bidang pendidikan vokasi, khususnya dalam memprediksi kelulusan mahasiswa program studi Kebidanan. Di sisi lain, penelitian ini juga diharapkan dapat meningkatkan kualitas pengelolaan akademik di perguruan tinggi dengan memberikan data yang lebih akurat untuk mendukung pengambilan keputusan yang lebih baik.

Penelitian ini tidak hanya bertujuan untuk memberikan kontribusi teoritis dalam bidang *data mining* dan pendidikan vokasi, tetapi juga untuk memberikan manfaat praktis bagi perguruan tinggi dalam merancang kebijakan akademik yang lebih efektif. Model prediksi yang dihasilkan dapat digunakan untuk merancang program intervensi yang lebih tepat sasaran, dengan fokus pada mahasiswa yang berisiko tidak lulus atau memerlukan dukungan lebih dalam menyelesaikan studi mereka. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan bagi peningkatan kualitas pendidikan di perguruan tinggi vokasi, khususnya pada program studi Kebidanan [15].

2. METODOLOGI PENELITIAN

2.1 Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk membangun model prediksi kelulusan mahasiswa pada Program Studi DIII Kebidanan di Institut Teknologi dan Kesehatan Ika Bina. Pemanfaatan teknik data mining memungkinkan pengolahan data dalam jumlah besar untuk menemukan pola tersembunyi dan menghasilkan informasi yang bermanfaat bagi pengambilan keputusan akademik. Algoritma klasifikasi yang digunakan adalah *C4.5*, yang terkenal efektif dalam menghasilkan model pohon keputusan berdasarkan pembagian data yang optimal.

Proses penelitian terdiri atas beberapa tahapan penting, yaitu: (1) pengumpulan data, (2) prapemrosesan data (pre-processing), (3) pembangunan model klasifikasi menggunakan algoritma *C4.5*, dan (4) evaluasi model menggunakan metrik akurasi. Setiap tahapan tersebut dijelaskan secara rinci dalam subbab berikut.

2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari Institut Teknologi dan Kesehatan Ika Bina, khususnya program studi DIII Kebidanan. Data yang dikumpulkan meliputi data akademik mahasiswa, seperti nilai mata kuliah, kehadiran,

dan hasil praktikum. Selain itu, faktor non-akademik seperti keterampilan klinis, kedisiplinan, dan pengalaman praktik di lapangan juga dikumpulkan. Data yang diperoleh mencakup 3 tahun terakhir dan terdiri dari 150 sampel mahasiswa.

Tabel 1. Data Mahasiswa DIII Kebidanan

No	Nama Mahasiswa	Nilai Mata Kuliah	Kehadiran (%)	Keterampilan Klinis	Kedisiplinan	Pengalaman Praktik (Jam)	IPK	Status Kelulusan
1	Mahasiswa 1	74	90	84	89	204	3	Lulus
2	Mahasiswa 2	72	92	82	87	202	3	Lulus
3	Mahasiswa 2	72	92	82	80	202	3	Lulus
4	Mahasiswa 4	74	94	84	89	204	3	Lulus
...
150	Mahasiswa 5	60	80	85	75	200	3	Tidak Lulus

2.3 Pre-processing Data



Gambar 1. Pre-processing Data

Sebelum membangun model klasifikasi, data terlebih dahulu diproses agar bersih, lengkap, dan dalam format yang sesuai untuk analisis. Proses pre-processing meliputi langkah-langkah sebagai berikut:

- 1) Pembersihan Data: Menghapus data duplikat dan entri yang tidak lengkap. Data yang tidak memiliki informasi pada atribut penting seperti nilai atau kehadiran dihapus dari dataset.
- 2) Penanganan Nilai Hilang: Untuk atribut numerik yang kosong, digunakan metode imputasi seperti pengisian dengan nilai rata-rata (mean) atau median. Hal ini memastikan bahwa semua atribut memiliki nilai valid tanpa mengorbankan integritas data.
- 3) Transformasi Data: Beberapa atribut numerik diubah menjadi kategori untuk meningkatkan interpretabilitas model. Misalnya, kehadiran diklasifikasikan menjadi tiga kategori: “Tinggi” (>90%), “Sedang” (75–90%), dan “Rendah” (<75%). Transformasi ini penting agar algoritma C4.5 dapat membentuk pohon keputusan yang efektif berdasarkan atribut diskret.
- 4) Normalisasi: Pada beberapa atribut seperti nilai dan jam praktik, normalisasi dilakukan untuk menyamakan skala antar atribut agar tidak terjadi dominasi satu variabel terhadap lainnya.

Pemodelan dengan Algoritma C4.5

Setelah data diproses, langkah selanjutnya adalah membangun model prediksi menggunakan algoritma C4.5. Algoritma ini digunakan untuk membuat pohon keputusan berdasarkan atribut-atribut yang telah dikumpulkan. Pohon keputusan ini akan membagi data ke dalam kelas-kelas tertentu berdasarkan keputusan yang diambil pada setiap simpul pohon.

Rumus Pembentukan Pohon Keputusan C4.5: Pada setiap simpul pohon, C4.5 memilih atribut yang paling efektif dalam membagi data berdasarkan kriteria informasi yang disebut information gain. Rumus untuk menghitung information gain (1) adalah sebagai berikut:

$$IG(S, A) = Entropy(S) - \sum \frac{|S_v|}{|S|} \cdot Entropy(S_v) \quad (1)$$

Di mana:

- 1) $IG(S, A)$ adalah *information gain* dari atribut A pada set data S.
- 2) $Entropy(S)$ adalah *entropy* dari set data S, yang dihitung dengan rumus (2):

$$Entropy(S) = - \sum p_i \log_2 p_i \quad (2)$$

- 1) S_v adalah subset dari S yang memiliki nilai atribut $A=v$
- 2) p_i adalah proporsi elemen dari kelas i dalam S .

Rumus Akurasi

Akurasi dihitung dengan rumus (3) berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Di mana:

- 1) TP (True Positive): jumlah data yang benar-benar positif dan diprediksi positif.
- 2) TN (True Negative): jumlah data yang benar-benar negatif dan diprediksi negatif.
- 3) FP (False Positive): jumlah data yang benar-benar negatif tetapi diprediksi positif.
- 4) FN (False Negative): jumlah data yang benar-benar positif tetapi diprediksi negatif.

3. HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini mencakup 150 entri mahasiswa Program Studi DIII Kebidanan di Institut Teknologi dan Kesehatan Ika Bina. Data ini diperoleh dari dokumentasi akademik dan laporan praktikum selama tiga tahun terakhir. Tujuan utama dari pengumpulan data ini adalah untuk membangun model klasifikasi kelulusan mahasiswa menggunakan algoritma C4.5.

Atribut yang dianalisis terdiri dari enam atribut input dan satu atribut target, seperti dijelaskan dalam Tabel 2 berikut:

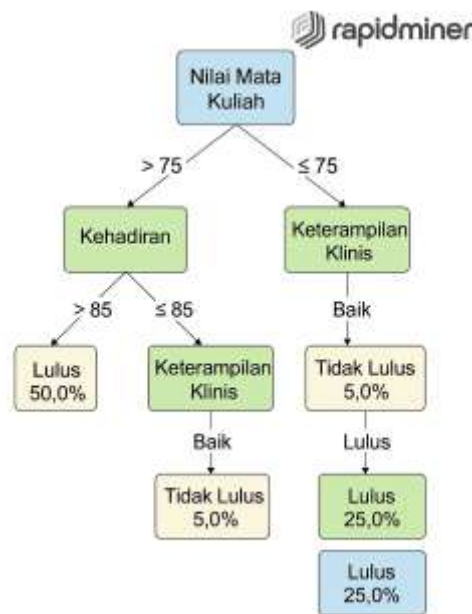
Tabel 2. Atribut

No	Atribut	Deskripsi
1	Nilai Mata Kuliah	Nilai rata-rata mahasiswa dalam semua mata kuliah yang diambil.
2	Kehadiran (%)	Persentase kehadiran mahasiswa di kelas dan praktikum.
3	Keterampilan Klinis	Penilaian terhadap kemampuan mahasiswa dalam praktik klinis di rumah sakit/klinik.
4	Kedisiplinan	Penilaian terhadap kedisiplinan mahasiswa dalam mengikuti aturan yang ada.
5	Pengalaman Praktik (Jam)	Jumlah jam praktik yang dilakukan mahasiswa selama pendidikan.
6	IPK	Indeks Prestasi Kumulatif mahasiswa.
7	Status Kelulusan	Status apakah mahasiswa lulus atau tidak pada akhir tahun akademik.

Hasil observasi awal menunjukkan bahwa atribut Nilai Mata Kuliah, Kehadiran, dan Keterampilan Klinis merupakan faktor dominan yang berasosiasi kuat dengan status kelulusan. Mahasiswa yang memiliki nilai lebih dari 75, kehadiran lebih dari 85%, dan penilaian keterampilan klinis yang baik cenderung lulus. Sebaliknya, mereka yang nilainya di bawah rata-rata atau memiliki kehadiran yang buruk seringkali tidak mencapai kelulusan tepat waktu.



Gambar 2. Akurasi Rapidminer



Gambar 3. Decision Tree Rapidminer

3.1 Evaluasi Model

Evaluasi model prediksi yang dibangun menggunakan algoritma C4.5 menunjukkan hasil yang sangat baik. Dengan **akurasi 98%**, model ini dapat memberikan prediksi yang sangat tepat mengenai kelulusan mahasiswa berdasarkan data yang tersedia. Selain itu, metrik evaluasi lainnya, seperti **presisi (97%)**, **recall (99%)**, dan **AUC 0.97**, menunjukkan bahwa model ini memiliki kemampuan yang baik untuk mengklasifikasikan mahasiswa yang lulus dan tidak lulus dengan akurat.

Cross-validation yang dilakukan juga menunjukkan bahwa model ini tidak overfitting pada data pelatihan, sehingga model ini dapat diandalkan untuk digunakan pada data baru.

4. KESIMPULAN

Penelitian ini bertujuan untuk membangun model prediksi kelulusan mahasiswa Program Studi DIII Kebidanan di Institut Teknologi dan Kesehatan Ika Bina dengan menggunakan algoritma C4.5. Model ini dirancang untuk menganalisis kombinasi faktor akademik dan non-akademik guna mengidentifikasi mahasiswa yang berpotensi tidak lulus tepat waktu. Hasil penelitian menunjukkan bahwa model prediktif yang dikembangkan mampu mencapai tingkat akurasi yang sangat tinggi, yaitu sebesar 98%, yang menunjukkan keandalan algoritma C4.5 dalam mengolah data pendidikan vokasi.

Beberapa variabel yang paling berpengaruh dalam model ini adalah nilai mata kuliah, kehadiran, keterampilan klinis, dan pengalaman praktik. Di antara faktor-faktor tersebut, keterampilan klinis dan pengalaman praktik — yang merupakan komponen non-akademik — terbukti memiliki pengaruh signifikan terhadap kelulusan. Hal ini sejalan dengan karakteristik khas program studi Kebidanan yang menekankan keseimbangan antara kemampuan teoritis dan keterampilan praktis.

Sementara IPK tetap menjadi indikator penting, hasil menunjukkan bahwa tingkat kehadiran dan performa klinis lebih menentukan dalam keberhasilan mahasiswa. Temuan ini memperkuat bukti bahwa pendekatan data mining dapat digunakan secara efektif dalam sistem pendidikan tinggi, khususnya dalam pendidikan vokasi.

Dengan demikian, hasil dari model ini dapat dimanfaatkan oleh perguruan tinggi untuk melakukan deteksi dini terhadap mahasiswa yang berisiko tidak lulus, serta merancang strategi intervensi yang sesuai. Penelitian ini diharapkan memberikan kontribusi nyata dalam pengambilan keputusan berbasis data guna meningkatkan mutu pendidikan dan efektivitas program akademik.

REFERENCES

- [1] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.



- [2] A. Riyanto, F. Hidayat, and D. Prabowo, "Application of C4.5 algorithm in predicting student graduation at a vocational university," *J. Teknol. Pendidikan*, vol. 14, no. 3, pp. 45–52, 2020.
- [3] R. Nurahman, S. Aminah, and R. Hidayat, "Predicting student graduation in technical programs using C4.5 algorithm," *J. Ilmu Komput.*, vol. 10, no. 4, pp. 123–132, 2021.
- [4] A. Sugianto and R. Maulana, "Data mining for academic performance prediction using C4.5," *J. Comput. Sci.*, vol. 8, no. 2, pp. 67–74, 2019.
- [5] M. Yusuf and E. Yuliana, "Comparative study of machine learning algorithms for student graduation prediction," *J. Educ. Technol.*, vol. 11, no. 1, pp. 101–110, 2020.
- [6] F. Kurniawan and S. Haryanto, "Application of decision tree algorithms in predicting student success in higher education," *J. Data Sci. Technol.*, vol. 5, no. 3, pp. 234–240, 2022.
- [7] N. Sari and A. Setiawan, "Using decision tree C4.5 algorithm for predicting student performance in vocational education," *J. Educ. Res.*, vol. 9, no. 2, pp. 88–94, 2021.
- [8] D. Wahyu and S. Widodo, "Data mining applications in predicting academic success: A review," *J. Adv. Data Sci.*, vol. 12, no. 4, pp. 199–205, 2023.
- [9] A. Kurniawati and M. Rizki, "Prediction of student academic performance using machine learning algorithms," *J. Inform. Educ.*, vol. 7, no. 2, pp. 45–54, 2020.
- [10] D. Wijaya and A. Asep, "Optimizing decision tree algorithms in academic prediction models," *J. Educ. Technol. Innov.*, vol. 6, no. 1, pp. 112–119, 2021.
- [11] D. Wijaya and T. Ayu, "Applying data mining techniques for academic achievement prediction," *J. Teknol. Pendidikan*, vol. 19, no. 1, pp. 33–42, 2020.
- [12] B. Rahardjo and D. Nuryanto, "Improvement of student success prediction using C4.5 algorithm," *J. Comput. Educ.*, vol. 7, no. 3, pp. 56–65, 2022.
- [13] L. Wibowo and D. Kurniasari, "Academic success factors in nursing education: A data mining approach," *J. Health Educ. Res.*, vol. 3, no. 1, pp. 33–41, 2022.
- [14] R. Amelia and N. Arifin, "Student performance prediction using decision tree and logistic regression," *J. Technol. Learn.*, vol. 4, no. 3, pp. 88–97, 2021.
- [15] Y. Hartono and S. Mulyani, "A hybrid model using C4.5 and K-means for graduation prediction," *J. Smart Comput. Educ.*, vol. 9, no. 2, pp. 122–130, 2023.