

Penerapan Decision Tree Pada Klasifikasi Kelulusan Mahasiswa

Zaehol Fatah¹, Muflihatul Hasanah^{2*}

^{1,2} Sistem Informasi, Universitas Ibrahimy, Situbondo

Email: ¹zaeholfatah@gmail.com, ^{2*}mflihatul03@gmail.com

Abstrak

Keberhasilan suatu program pendidikan sering diukur dari jumlah mahasiswa yang menyelesaikan studinya tepat waktu, terutama dalam hal evaluasi dan akreditasi universitas. Penelitian ini bertujuan untuk penerapan algoritma Pohon Keputusan dalam proses prediksi terhadap kelulusan mahasiswa tepat waktu. Metode tersebut digunakan karena mampu mengenali pola data secara terstruktur dan memberikan hasil prediksi yang akurat. Dalam studi ini, data dikumpulkan dari total 154 mahasiswa, dengan atribut utama meliputi IPK, IPS, umur, dan jenis kelamin. Proses pengolahan data dilakukan melalui tahapan seleksi data, pembersihan, transformasi, dan pemodelan menggunakan perangkat lunak RapidMiner. Model kemudian dievaluasi dengan metode cross-validation untuk mencegah overfitting serta memastikan konsistensi performa model. Hasil pengujian memperlihatkan bahwa algoritma Pohon Keputusan efektif dalam pengelompokan data kelulusan mahasiswa dengan akurasi yang tinggi sehingga mencapai 94,48%. Hal ini menegaskan bahwa metode ini dapat diandalkan sebagai alat bantu prediktif bagi perguruan tinggi dalam memantau potensi kelulusan mahasiswa sejak dini. Dengan demikian, implementasi model klasifikasi ini diharapkan dapat mendukung pengambilan keputusan dalam perencanaan akademik serta meningkatkan efisiensi manajemen pendidikan di perguruan tinggi.

Kata Kunci: *Prediksi kelulusan mahasiswa, Sistem Informasi, Algoritma Decision Tree, Dataset Kelulusan.*

PENDAHULUAN

Pendidikan berperan krusial dalam membentuk karakter dan perkembangan individu masyarakat serta membangun negara, sehingga menjadi unsur utama dalam mencapai tujuan pembangunan. Salah satu indikator penting untuk mengevaluasi keberhasilan sistem pendidikan adalah tingkat kelulusan mahasiswa secara tepat waktu (Putri & Wijayanto, 2022).

Universitas merupakan lembaga pendidikan yang bertugas menyelenggarakan kegiatan pembelajaran dan mengabdikan kepada masyarakat, yang dikenal sebagai bagian dari Tiga Pilar Pendidikan Tinggi (Zeniarta et al., 2022). Gelar sarjana hanya dapat diperoleh mahasiswa yang berhasil menyelesaikan studinya tepat waktu (Nugroho & Wibowo, 2017). Mahasiswa dapat memperoleh status kelulusan tepat waktu jika mereka menyelesaikan pendidikan sarjana mereka,

yang berarti studi selesai dalam jangka waktu empat tahun atau bahkan lebih cepat.

Salah satu cara untuk membuat prediksi adalah dengan menggunakan teknik yang disebut penambangan data, yang melibatkan penjelajahan data untuk menemukan informasi yang berguna (Novianto et al., 2023). Pohon Keputusan cocok untuk memprediksi kelulusan mahasiswa secara akurat karena fleksibel dalam memilih fitur, yang mengarah ke simpul internal yang berbeda. Fitur-fitur yang dipilih akan memengaruhi satu sama lain dalam node yang sama, menyebabkan kebingungan dan frustrasi ketika fitur-fitur tersebut tidak selaras, yang menyebabkan node internal yang berbeda (Permatasari, 2021). Dalam beberapa tahun terakhir, berbagai penelitian telah memanfaatkan metode Decision Tree untuk mengklasifikasikan kelulusan mahasiswa. Metode ini mampu menghasilkan tingkat akurasi yang tinggi,

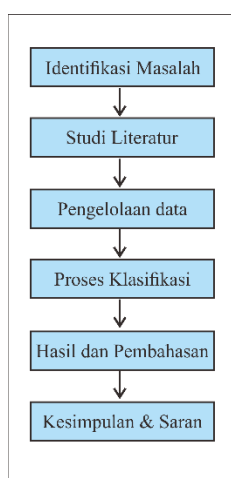
sehingga berpotensi menjadi alternatif solusi dalam membantu meningkatkan keberhasilan mahasiswa dalam menyelesaikan pendidikan (Qisthiano et al., 2023).

Untuk melakukan prediksi mahasiswa, faktor nilai dan ekonomi merupakan faktor utama dalam penentuan atribut, dari segi nilai digunakan atribut IPK, dan Setelah menentukan atribut yang paling utama dan yang paling penting, selanjutnya memilih atribut duluar faktor nilai yaitu jenis kelamin

Perguruan Tinggi saat ini semakin intens, perguruan tinggi dituntut untuk senantiasa melakukan perbaikan manajerial guna meningkatkan kualitas pembelajaran dan pening katan kualitas pembelajaran dan peringkat akreditasi. Penilaian akreditasi sendidi turut mempertimbangkan aspek mahasiswa dan lulusan.

METODE

Rancangan penelitian ini melibatkan sejumlah langkah yang harus dilalui dengan menggunakan pendekatan decision tree dan Naive Bayes diterapkan untuk mengelompokkan data dan memperkirakan kelulusan mahasiswa secara tepat waktu. Dalam proses pengolahan data menggunakan *RapidMiner* sebagai alat bantu. Rangkaian tahapan bisa dilihat di gambar 1.1 beriku:



Gambar 1. Tahapan penelitian

1. Langkah awal dalam penelitian ini adalah mengidentifikasi serta

merumuskan permasalahan yang akan dikaji

2. Tahapan berikutnya adalah melakukan kajian literatur untuk memahami berbagai penelitian terdahulu yang relevan, Khususnya mengenai prediksi tepat waktu data kelulusan mahasiswa dengan memanfaatkan algoritma *Decision Tree* atau Algoritma Pohon Keputusan
3. Melaksanakan proses pengolahan data, dimulai dari proses sleksi data, pembersihan, transformasi, hingga data sisap digunakan dalam pemodelan, menggunakan Rapid Miner
4. Proses pengujian algoritma pohon keputusan terbagi atas dua bagian : data pelatihan dan data uji , serta pelaksanaan pengujian itu sendiri.
5. Tahap analisis hasil proses pembahasan dari hasil yang diperoleh dimana fokusnya adalah pada tingkat akurasi dalam mengklasifikasi data prediksi kelulusan mahasiswa tepat waktu
6. Terahir, tahapan penarikan kesimpulan, peneliti menyampaikan hasil akhir dari peneliti menyampaikan hasil akhir dari penelitian berdasarkan temuan dan analisis yang telah dilakukan (Qisthiano et al., 2023)

Data Penelitian

Dalam pelaksanaan penelitian, untuk mendukung proses analisis, dibutuhkan data publik sebagai acuan, yang dalam hal ini berupa data sampel kelulusan dari 145 mahasiswa. Variabel-variabel yang dianalisis dalam penelitian ini meliputi umur, IPS dan IPK. Data ini digunakan untuk mendukung proses klasifikasi prediksi kelulusan mahasiswa dengan pendekatan algoritma *Decision Tree*.

Pengumpulan Data

Metode diterapkan oleh peneliti ini terdiri dari beberapa tahapan sebagai berikut

1. Melakukan studi literatur guna mamperoleh landasan teori serta referensi dari peneliti terdahulu yang relefan dengan topik penelitian ini.

2. Pengumpulan data dilakukan dengan mengakses data, yang kemudian dijadikan sebagai dasar acuan dalam menganalisis data mahasiswa aktif, dengan mempertimbangkan kesesuaian terhadap atribut-atribut yang digunakan didalam penelitian(Iqbal et al., 2024)

Klasifikasi

Klasifikasi merupakan metode untuk membentuk model yang merepresentasikan karakteristik setiap kelas, agar dapat mengidentifikasi kelas dari data baru yang belum diketahui kategorinya(Menggunakan & Metode, 2020).

Data Mining

Disebut data mining karena metode yang memungkinkan penggunaanya untuk mengakses data yang rentang lebih besar dalam waktu yang relatif cepat. Data ming merupakan proses perspektif yang bervariasi lalu diolah menjadi informasi

yang bernilai guna(Septian, 2009). Proses Data Mining dilakukan dengan memanfaatkan data yang tersedia untuk membentuk suatu model, yang kemudian digunakan dalam mengenali pola dari data lain diluar data yang tersimpan(Qisthiano et al., 2023). Data mining adalah proses untuk mengidentifikasi pola-pola yang bernilai dari kumpulan data berukuran besar, yang biasanya tersimpan dalam database, data warehouse, atau media penyimpanan(Lestari, 2017).

HASIL DAN PEMBAHASAN

Hasil Pengujian Rapidminer

Proses pengujian dilakukan menggunakan perangkat lunak RapidMiner guna mengevaluasi akurasi algoritma Pohon Keputusan dalam memprediksi kelulusan mahasiswa yang ada pada Tabel 1.

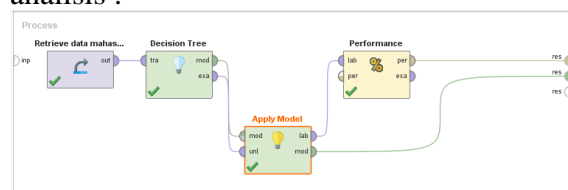
Tabel 1. Dataset

NAMA	UMUR	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK	STATUS KELULUSAN
FANDI AHMAD	22	3,02	3,42	3,81	3,61	3,35	3,43	3,46	4	3,49	TEPAT
ALI NURSID	22	2,6	2,2	2,27	2,55	2,45	2,39	2,55	0,75	2,7	UJIAN ULANG
DIDI MULYANTO	22	2,43	2,55	2,64	2,68	2,68	2,68	2,67	3	2,9	UJIAN ULANG
KHOIRUL ULUM	22	2,52	2,7	2,73	3,27	2,77	2,53	3,25	3	2,99	UJIAN ULANG
DEDY DANU											
SAPUTRA	22	2,86	2,86	3,25	3,25	3,31	3,48	2,43	3	3,28	TEPAT
DERA BAHTIAR											
WIDIYANTO	23	2,52	3,3	3,79	3,71	3,65	2,1	3,59	4	3,52	TEPAT
ULIL ALBAB	23	2,93	3,18	3,31	3,25	3,44	3,22	3,07	3,5	3,33	TEPAT
AGUS NUGROHO	23	2,83	3,11	3,1	2,96	3,2	3,53	2,48	3	3,27	TEPAT
AWIT RIKA JAYA											
KUSUMA	23	2,74	2,86	2,61	2,95	2,86	2,68	3,16	2,5	2,92	TEPAT
FACHRY											
WICAKSONO	23	2,83	3	3,55	2,82	3,14	2,73	3,3	3	3,17	TEPAT
ANDI PRASETYO	23	2,71	2,41	2,8	2,95	3,18	3,1	2,5	3	3,03	TEPAT
NOOR YANTO	23	2,98	2,43	2,52	2,86	2,2	2,83	2,73	0	2,62	UJIAN ULANG
FAHRUR ABID	23	2,62	2,43	2,11	2,32	2,48	2,33	1,43	1,5	2,53	UJIAN ULANG
DINA FITHRONI	23	2,6	2,52	2,3	2,77	2,55	2,48	2	4	2,52	UJIAN ULANG
AGUSTINUS	23	2,98	2,2	2,16	1,8	2,11	1	1,19	0,79	2,12	UJIAN ULANG
SHOFIL GUNAWAN	23	2,6	1,93	2,37	2,09	2,41	2,36	1,26	1,6	2,41	UJIAN ULANG
JAMAL HENDRI											
PRASETIYO	23	3,31	2,96	3,48	3,26	3,15	3,33	3,68	0	3,32	TEPAT
AUFAL WIDAD	23	2,5	2,08	2,25	2,27	2,21	2,16	2,86	2,89	2,8	UJIAN ULANG
EKO SUHERMAN	23	1,79	2,24	2,02	2,45	2,36	2,43	2,8	2,8	2,72	UJIAN ULANG
ROBERT MYLLANO	23	2,26	2,61	2,34	2,86	2,23	2,75	2,39	3,69	2,75	UJIAN ULANG
ABDUL LATIF	23	2,55	2,75	2,64	2,59	2,8	2,75	1,14	0	2,46	UJIAN ULANG
AHMAD FAIZIN	23	3	2,89	3,25	2,83	2,86	2,97	1,31	0	2,88	UJIAN ULANG

Proses persiapan data di RapidMiner menunjukkan bagaimana data disortir dan diatur untuk analisis, termasuk menghapus data yang tidak relevan, menangani nilai

yang hilang, dan menyesuaikan format data menurut tipe data dan variabel yang ditentukan, yang membantu menghindari

kebingungan dan frustrasi selama proses analisis .



Gambar 2. Validation Decision Tree

Gambar ini menunjukkan alur proses dalam suatu perangkat lunak data mining. Alur kerja dimulai dengan operator Retrieve data mahasiswa yang berfungsi mengambil dataset awal. Dataset ini kemudian dihubungkan ke operator Decision Tree, yang bertugas melatih model pohon keputusan menggunakan data pelatihan. Kemudian model yang telah dilatih dan data yang diberi label selanjutnya dimasukkan ke operator Apply Model. Operator ini menggunakan model yang sudah dilatih untuk menghasilkan prediksi. Output dari Apply Model diumpun ke Performance, yang mengevaluasi kualitas prediksi model dengan menghitung metrik performa berdasarkan data berlabel yang dihasilkan (Mustapha et al., 2023)

Dengan menerapkan teknik Dengan menggunakan Validasi Silang , kami memperoleh hasil evaluasi kinerja model yang lebih andal , seperti akurasi, presisi, dan penarikan kembali. Validasi Silang dalam algoritma Pohon Keputusan berfungsi sebagai metode evaluasi Untuk mengevaluasi keakuratan model , kumpulan data dibagi menjadi beberapa subset yang disebut lipatan. Model kemudian dilatih menggunakan bagian dari subset dan diuji pada bagian lain . Pendekatan ini membantu menurunkan risiko overfitting dan memberikan gambaran yang lebih stabil dan akurat tentang kinerja model (Putra & Yptk, 2025).

Tree

```

IPK > 2.995: TEPAT (TEPAT=77, UJIAN ULANG=1, TERLAMBAT=0)
IPK ≤ 2.995
| IPK > 2.365
| | IPK > 2.400: UJIAN ULANG (TEPAT=1, UJIAN ULANG=50, TERLAMBAT=3)
| | IPK ≤ 2.400: TEPAT (TEPAT=1, UJIAN ULANG=1, TERLAMBAT=0)
| IPK ≤ 2.365: TERLAMBAT (TEPAT=0, UJIAN ULANG=2, TERLAMBAT=9)
    
```

Gambar 3. Hasil Dari Decision Tree

Gambar di atas menampilkan hasil Visualisasi pohon keputusan dibuat melalui proses klasifikasi pada data siswa menggunakan algoritma Pohon Keputusan di RapidMiner. Pohon keputusan ini digunakan untuk memprediksi kelulusan mahasiswa berdasarkan atribut Indeks Prestasi Kumulatif (IPK). Pohon keputusan bekerja dengan membagi data menjadi simpul - simpul berdasarkan nilai ambang atribut IPK, kemudian menetapkan kelas hasil seperti "TEPAT", "UJIAN ULANG", atau "TERLAMBAT" pada setiap cabangnya. Proses klasifikasi dimulai dari simpul akar (root node) yang memisahkan data berdasarkan nilai $IPK > 2.995$. Mahasiswa dengan IPK di atas nilai ini langsung diprediksi lulus tepat waktu, yang ditunjukkan dengan label TEPAT (jumlah: 77), sementara sebagian kecil berada pada kategori UJIAN ULANG (1 mahasiswa), dan tidak ada yang TERLAMBAT. Selanjutnya, bagi mahasiswa dengan $IPK \leq 2.995$, dilakukan pemisahan lebih lanjut berdasarkan ambang batas IPK 2.365. Jika $IPK > 2.365$, maka terdapat percabangan baru pada nilai 2.400. Mahasiswa dengan $IPK > 2.400$ umumnya berada pada kategori UJIAN ULANG, sedangkan yang IPK-nya antara 2.365 hingga 2.400 diprediksi tetap TEPAT, meskipun hanya sejumlah kecil. Sementara itu, mahasiswa dengan $IPK \leq 2.365$ diklasifikasikan sebagai TERLAMBAT, yang menunjukkan bahwa semakin rendah IPK, maka kemungkinan mahasiswa mengalami keterlambatan kelulusan semakin besar. Model ini mencerminkan proses pengambilan keputusan berjenjang yang menjadi ciri khas dari algoritma Decision Tree, Setiap simpul internal dalam pohon keputusan menggambarkan Proses pengujian untuk atribut tertentu , dan setiap

cabang menunjukkan hasil dari proses pengujian. Proses ini berlanjut hingga mencapai simpul daun (leaf node) yang menunjukkan hasil prediksi. Hasil visualisasi ini sangat membantu dalam memahami pola distribusi kelulusan mahasiswa berdasarkan IPK, dan mendukung tujuan penelitian untuk mengidentifikasi faktor akademik yang paling memengaruhi ketepatan waktu kelulusan. Hal ini juga sesuai dengan pernyataan dalam jurnal bahwa Decision Tree memiliki fleksibilitas tinggi dalam memilih fitur yang relevan, serta memberikan interpretasi hasil yang mudah dipahami oleh pengguna non-teknis

accuracy: 94.48%

	true TEPAT	true UJIAN ULANG	true TERLAMBAT	class precision
pred. TEPAT	78	2	0	97.50%
pred. UJIAN ULANG	1	78	3	92.59%
pred. TERLAMBAT	0	2	9	81.82%
class recall	98.73%	92.59%	75.00%	

Gambar 4. Hasil Akurasi Menggunakan Decision Tree

SIMPULAN (PENUTUP)

Studi ini meneliti penggunaan algoritma Pohon Keputusan untuk memprediksi peluang kelulusan siswa. Studi ini memiliki urgensi tinggi karena tepatnya kelulusan merupakan salah satu tolak ukur kinerja utama yang berkontribusi terhadap penilaian mutu lembaga pendidikan. Sebagai bagian dari upaya prediktif menuju kelulusan , data yang digunakan berasal dari platform Kaggle yang terdiri dari 154 data siswa .

Kelebihan algoritma Decision Tree terbukti mampu memetakan pola data secara efektif dengan tingkat akurasi yang tinggi. Evaluasi performa model dilakukan menggunakan sejumlah metrik evaluasi serta validasi tambahan guna memastikan keakuratan hasil prediksi. Penelitian ini berhasil menciptakan model klasifikasi yang andal untuk memprediksi kelulusan siswa . Model ini dibangun menggunakan algoritma Pohon Keputusan dan diuji dengan 20 % data sebagai set uji dari seluruh kumpulan data . Hasilnya menunjukkan kinerja yang sangat baik ,

dengan model mencapai akurasi sebesar 94,48 %.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih sebesar-besarnya kepada Dosen Pembimbing atas bimbingan dan masukan krusial yang diberikan dalam penelitian ini. Dukungan yang tiada henti dari keluarga, sahabat, dan teman dekat turut menjadi kunci kelancaran proses ini. Besar harapan penulis, kiranya hasil penelitian ini dapat memberikan manfaat signifikan bagi pembaca.

DAFTAR PUSTAKA

Iqbal, M., Donny, M., Wahyu, H., Tedi, K. N., Amali, A., & Ismasari, N. (2024). Analysis of K-Nearest Neighbor (KNN), Naive Bayes ands Decision Tree C4.5 Algorithm With Classification Method In Breast Cancer Using RapidMiner. *Journal of Applied Intelligent System*, 9(2), 260–270. <https://doi.org/10.62411/jais.v9i2.11242>

Lestari, N. A. (2017). *PENERAPAN DATA MINING MENGGUNAKAN METODE DECISION TREE C 4.5 UNTUK PREDIKSI TINGKAT KELULUSAN MAHASISWA (STUDI KASUS : STM IK WIT)*.

Menggunakan, P., & Metode, C. (2020). *Prediksi Kelulusan Mahasiswa Berdasarkan Data Berkunjung*. 8(2), 24–33.

Mustapha, M. F., Zulkifli, A. N. I., Kairan, O., Zizi, N. N. S. M., Yahya, N. N., & Mohamad, N. M. (2023). The prediction of student’s academic performance using RapidMiner. *Indonesian Journal of Electrical Engineering and Computer Science*, 32(1), 363–371. <https://doi.org/10.11591/ijeecs.v32.i1.pp363-371>

Novianto, E., Hermawan, A., & Avianto, D. (2023). *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 8(2), 146–

154.

<https://doi.org/10.36341/rabit.v8i2.34>

34

- Nugroho, M. F., & Wibowo, S. (2017). Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes. *Jurnal Informatika Upgris*, 3(1), 63–70. <https://doi.org/10.26877/jiu.v3i1.1669>
- Permatasari, R. P. (2021). *Implementasi algoritma decision tree untuk prediksi kelulusan mahasiswa tepat waktu laporan skripsi*.
- Putra, U., & Yptk, I. (2025). Penerapan algoritma klasifikasi untuk prediksi tingkat kelulusan mahasiswa menggunakan rapidminer. 6(1), 376–388. <https://doi.org/10.46576/djtechno>
- Putri, N. B., & Wijayanto, A. W. (2022). Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing. *Komputika : Jurnal Sistem Komputer*, 11(1), 59–66. <https://doi.org/10.34010/komputika.v11i1.4350>
- Qisthiano, M. R., Prayesy, P. A., & Ruswita, I. (2023). Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa. *G-Tech: Jurnal Teknologi Terapan*, 7(1), 21–28. <https://doi.org/10.33379/gtech.v7i1.1850>
- Septian, N. Y. (2009). Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro. *Jurnal Semantik 2013*, 1–11.
- Zeniarja, J., Salam, A., & Ma'ruf, F. A. (2022). Seleksi Fitur dan Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa. *Jurnal Rekayasa Elektrika*, 18(2), 102–108. <https://doi.org/10.17529/jre.v18i2.24047>