

Seleksi Fitur SelectKBest Dalam Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Decision Tree

Yogi Erka Julyansa Putra^{1*}, Imelda Imelda², Suryadih Suryadih¹

¹ Fakultas Teknologi Informasi, Magister Ilmu Komputer, Universitas Budi Luhur, Jakarta Selatan, Indonesia

² Departemen Teknologi Informasi, Magister Ilmu Komputer, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: ^{1*}2211601816@student.budiluhur.ac.id, ²imelda@budiluhur.ac.id, ³2211601873@student.budiluhur.ac.id

Email Penulis Korespondensi: 2211601816@student.budiluhur.ac.id

Submitted: 03/03/2025; Accepted: 26/03/2025; Published: 28/03/2025

Abstrak—Perguruan tinggi di Jakarta mengalami masalah dengan lonjakan jumlah mahasiswa yang tidak lulus tepat waktu di kantor organisasi. Rekomendasi ini menganalisis pertunjukan yang digunakan untuk memprediksi kelulusan mahasiswa tepat waktu. Tujuan utama dari pertimbangan ini adalah untuk membuat ilustrasi perkiraan dengan menggunakan Decision Tree. Data yang digunakan menggabungkan informasi tentang komponen-komponen yang berbeda. Evaluasi eksekusi yang muncul dilakukan berdasarkan estimasi seperti ketepatan, ketelitian, dan peninjauan. Kelulusan tepat waktu memberikan manfaat yang beragam bagi perguruan tinggi. Pertama-tama, hal ini meningkatkan reputasi institusi sebagai penyedia instruksi berkualitas yang mendukung mahasiswa untuk menambah pertimbangan mereka hampir tepat waktu. Perguruan tinggi dapat memajukan kualitas dan kelangsungan instruksi dengan mengaktualisasikan metode berdasarkan permintaan ini. Pertimbangan tersebut membuat aplikasi bantalan desktop untuk memasukkan data mahasiswa pengganti dan memperkirakan apakah setiap mahasiswa pengganti akan lulus tepat waktu atau tidak. Pemeriksaan ini membuat komitmen penting untuk upaya yang ditujukan untuk memajukan kualitas pengajaran di perguruan tinggi dan membuat para mahasiswa pengganti memenuhi tujuan wawasan mereka dengan lebih baik. Data diambil dari universitas di Jakarta terdiri dari 308 data mahasiswa. Hasil yang didapat hampir menggambarkan bahwa ilustrasi yang menggunakan jaminan sorotan menghasilkan akurasi 97,85%, sedangkan ilustrasi tanpa pilihan sorotan menghasilkan akurasi 93,54%.

Kata Kunci: Machine Learning; Prediksi; Klasifikasi; Mahasiswa; SciKit Learn

Abstract—University in Jakarta are facing issues with a surge in the number of students not graduating on time within the organizational office. This recommendation analyzes the performance used to predict timely student graduation. The primary objective of this consideration is to create an estimation illustration using a Decision Tree. The data used combines information about various components. The evaluation of the emerging execution is based on metrics such as accuracy, precision, and review. Timely graduation provides various benefits to colleges. Firstly, it enhances the institution's reputation as a provider of quality instruction that supports students in completing their studies almost on time. Colleges can improve the quality and sustainability of instruction by implementing methods based on this demand. The consideration involves creating a desktop application to input substitute student data and predict whether each substitute student will graduate on time or not. This examination makes a significant contribution to efforts aimed at advancing the quality of teaching in colleges and helping substitute students better achieve their academic goals. The data, collected from universities in Jakarta, consists of 308 student records. The results almost illustrate that the model using highlight assurance yields an accuracy of 97.85%, while the model without highlight options yields an accuracy of 93.54%.

Keywords: Machine Learning; Prediction; Classification; Student; Scikit Learn

1. PENDAHULUAN

Prediksi siswa tepat waktu dibuat merupakan aplikasi *desktop* yang memiliki kemampuan penambahan data, dan untuk memprediksi apakah mahasiswa lulus tepat waktu, teknik analisis data untuk mengidentifikasi pola dan hubungan dalam data akademik. Proses ini melibatkan pencatatan data historis seperti nilai, keberadaan, aktivitas ekstrakurikuler, faktor demografis, dan banyak lagi, diikuti oleh data seperti pohon, *Naïve Bayes multinomial*, *Naïve Bayes Guacian*, dan lebih banyak lagi untuk membangun model prediktif [1]. Menggunakan metode validasi seperti akurasi, *precision*, *recall*, dan *f1 score* sehingga lembaga tidak dapat dievaluasi dan dioptimalkan untuk model yang dikembangkan untuk ditutup dalam mengidentifikasi siswa yang tidak berisiko. Selesai pada waktu yang tepat dan mengambil tindakan pencegahan yang diperlukan [2]. Jumlah mahasiswa yang tidak dapat menyelesaikan lulus pada waktu yang tepat dan menjamin mutu akreditasi adalah tantangan serius bagi universitas. Jika ini tidak diatasi, ini dapat merusak stabilitas akreditasi, mengurangi daya saing, dan merusak lembaga pendidikan.

Pada tahun 2021, peningkatan mahasiswa lulus terlambat dan merupakan tantangan serius untuk memastikan kesimpulan kelulusan mahasiswa tidak terlambat. Penurunan mahasiswa tepat waktu dapat mempengaruhi nilai akreditasi [3], [4]. Jika gangguan ini tidak diperlakukan dengan serius, ia dapat memiliki dampak negatif yang berbeda pada lembaga. Jika ini diulang terus menerus, itu mengarah pada kegagalan proses akademik, terutama yang berkaitan dengan tujuan menghabiskan waktu, yang mengarah pada pembentukan data dari manajemen siswa, data akademik, dan kegiatan akademik lainnya yang terhambat. Biaya akhir yang rendah dapat menempatkan stabilitas keuangan dan akreditasi kelembagaan dalam risiko, dan bahkan dapat menyebabkan kebangkrutan. Selain itu, efek ini juga terasa pada tingkat daya saing kelembagaan dalam mencari siswa baru.

Banyak penelitian sebelumnya menggunakan metode klasifikasi spesifik dan mempertimbangkan berbagai atribut, seperti jenis kelamin dan latar belakang pendidikan. Namun, perlu untuk memperluas ruang lingkup atribut yang digunakan dan memperdalam pemahaman kita tentang faktor-faktor non-akademik yang mempengaruhi

kesimpulan periode itu. Meskipun beberapa penelitian sebelumnya telah berhasil menunjukkan efektivitas taksonomi spesifik dalam memprediksi lulus dalam waktu tepat, masih ada hambatan untuk memeriksa potensi keseluruhan untuk memprediksi kesimpulan pada waktu yang tepat. Akibatnya, kurangnya pemahaman komprehensif tentang faktor-faktor yang mempengaruhi kesimpulan dan pengembangan solusi yang efektif untuk meningkatkan penilaian akhir [5]. Teknik yang digunakan pada penelitian sebelumnya adalah *Decision Tree Classification (CART)*. Penelitian sebelumnya juga telah menciptakan model untuk memprediksi kelulusan tepat waktu siswa. Keterampilan pohon keputusan pada sebuah penelitian di Filipina mencapai akurasi yang tinggi yaitu 85,31% [2]. Menerapkan teknik pemilihan fitur agar atribut yang digunakan berkontribusi maksimal pada model. Decision Tree menghasilkan model yang dapat digambarkan secara visual, sehingga memudahkan pemahaman dan interpretasi, terutama bagi individu non-teknis atau pihak yang membuat keputusan [6], [7]. Pemilihan fitur adalah proses memilih subkumpulan atribut yang relevan untuk digunakan dalam model pembelajaran mesin. Teknik ini memungkinkan Anda menghapus atribut yang memiliki sedikit atau tidak ada kontribusi, meningkatkan performa model, mengurangi kompleksitas, dan menghindari *overfitting*. Pada penelitian ini diharapkan diperoleh hasil prediksi yang lebih akurat dan efisien dengan merealisasikan pemilihan fitur menggunakan *SelectKBest* dengan metode *f_classif* [8], [9]. Memperluas atribut meningkatkan akurasi karena semakin banyak informasi yang dimasukkan ke dalam model, semakin baik kemampuan model dalam mengenali pola dan hubungan kompleks dalam data [10], [11].

Analisis data akademik yang lebih dalam dapat mengidentifikasi siswa yang berisiko tidak menyelesaikan penelitian mereka pada waktu sebelum masalah muncul. Penggunaan keputusan teknik seperti *Decision Tree (CART)* dan algoritma lainnya dapat memberikan wawasan tentang bagaimana kombinasi faktor yang berbeda dan non-akademik mempengaruhi keberhasilan akademik siswa. Menggunakan data historis, model prediktif ini dapat menampilkan pola yang tidak selalu terlihat melalui pengamatan langsung. Ini memungkinkan lembaga untuk mengembangkan lebih banyak program intervensi yang ditargetkan, dan berdasarkan bukti berisiko untuk mengurangi ekstensi tingkat akhir dan meminimalkan dampak negatif pada reputasi kelembagaan [12].

Teknik ini telah terbukti efektif dalam beberapa penelitian sebelumnya. Pengembangan model yang dapat banyak digunakan di berbagai lembaga pendidikan tetap menjadi tantangan. Setiap universitas memiliki konteks yang berbeda, seperti struktur kurikulum, politik akademik, dan profil demografis siswa, dan setiap orang dapat mempengaruhi keakuratan model prediktif [13]. Oleh karena itu, pengujian lebih lanjut diperlukan untuk mengadaptasi model-model ini dengan kondisi tertentu untuk setiap Universitas [14]. Hal ini memungkinkan universitas pendidikan untuk mengambil langkah-langkah yang ditargetkan untuk membantu siswa dalam keberhasilan akademik, karena hasil kesimpulan gelar menjadi lebih relevan dan dapat diandalkan.

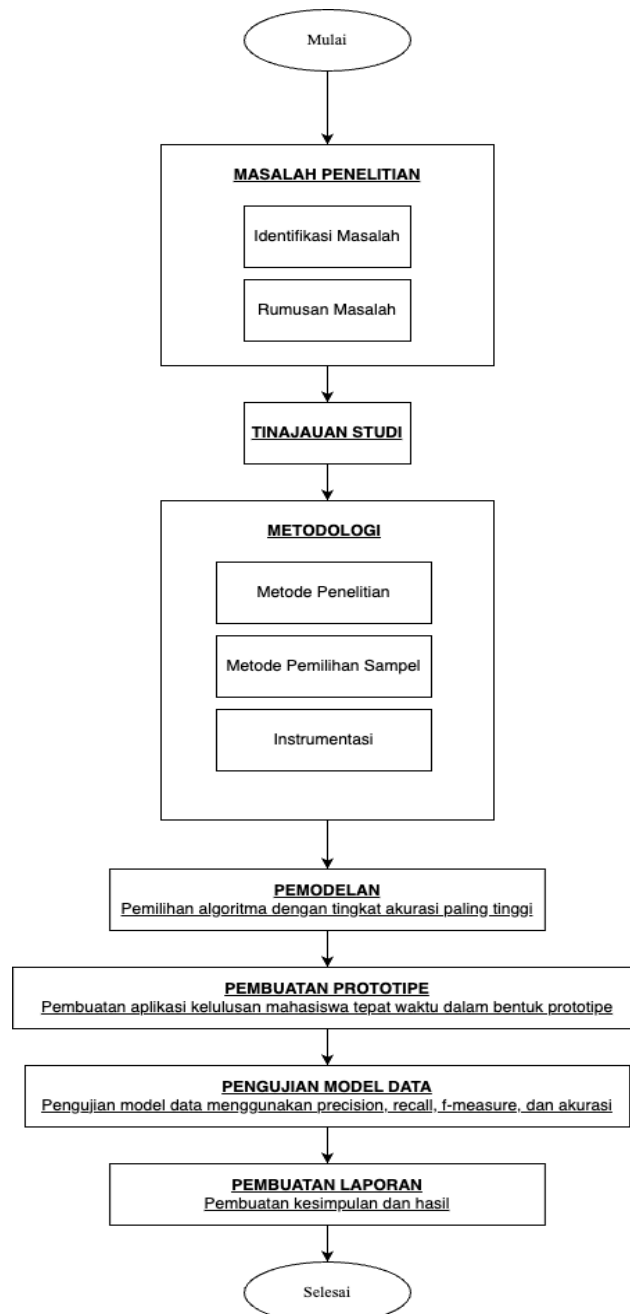
Penggunaan metode *SelectKBest* dalam proses pemilihan fitur memberikan dampak positif yang signifikan terhadap peningkatan kinerja model prediksi kelulusan tepat waktu. Teknik ini mengevaluasi hubungan antara masing-masing fitur dan variabel target, lalu memilih sejumlah k fitur dengan skor tertinggi. Dalam konteks prediksi kelulusan, fitur-fitur seperti IPK, IPS tiap semester, dan data demografis seperti usia memiliki tingkat pengaruh yang bervariasi. Dengan memanfaatkan fungsi *f_classif* untuk evaluasi, *SelectKBest* mampu mengidentifikasi fitur-fitur yang secara statistik paling signifikan, sehingga memungkinkan pengurangan jumlah fitur tanpa kehilangan informasi penting. Pendekatan ini tidak hanya mempercepat proses komputasi, tetapi juga mengurangi risiko *overfitting* yang umum terjadi pada model berdimensi tinggi, terutama saat jumlah data pelatihan terbatas dibandingkan banyaknya fitur.

Untuk mengukur performa model *Decision Tree* yang menggunakan *SelectKBest*, dilakukan validasi silang (*cross-validation*) guna memperoleh hasil yang lebih stabil dan dapat diterapkan secara umum. Melalui pengujian terhadap berbagai kombinasi nilai k pada *SelectKBest* serta kedalaman pohon pada algoritma *Decision Tree (CART)*, diperoleh konfigurasi optimal yang mampu menghasilkan model dengan keseimbangan antara akurasi, presisi, *recall*, dan *f1-score*. Hasil analisis lebih lanjut menunjukkan bahwa faktor akademik, khususnya IPK dan IPS pada semester awal, merupakan kontributor utama dalam memprediksi kelulusan tepat waktu. Temuan ini menekankan pentingnya pemberian intervensi sejak dini kepada mahasiswa yang mengalami penurunan prestasi akademik. Selain itu, visualisasi dari pohon keputusan yang terbentuk mempermudah pihak kampus dalam mengidentifikasi pola dan titik kritis dalam perjalanan studi mahasiswa yang dapat menyebabkan keterlambatan kelulusan.

Penelitian sebelumnya menunjukkan variasi hasil akurasi dengan berbagai metode klasifikasi. Hartatik et al. (2020) menggunakan Naïve Bayes dan mencapai akurasi 89% [2], sedangkan Gunawan et al. (2021) menggabungkan Naïve Bayes dengan KNN dan memperoleh hasil lebih tinggi, yaitu 90% [15]. Di sisi lain, Anwar et al. (2022) menerapkan Decision Tree dan C4.5 dengan akurasi tertinggi sebesar 92% [16], sementara Lutunani & Nugroho (2023) juga menggunakan Decision Tree dan mencapai 91,25% [17]. Sebagai perbandingan, Wirawan (2020) mengimplementasikan C4.5 dengan hasil 89,82% [18]. Temuan-temuan ini menunjukkan bahwa metode Decision Tree dan kombinasi algoritma cenderung memberikan akurasi lebih tinggi dibandingkan Naïve Bayes tunggal dalam beberapa kasus.

Berbeda dengan penelitian sebelumnya, solusi penelitian ini membutuhkan penggunaan metode pohon (*CART*) dan seleksi *selectKbest* untuk membuat keputusan untuk memprediksi kesimpulan pada waktu yang tepat. Memilih metode pohon keputusan berdasarkan, dan akurasinya adalah 85,31%. Pilihan fitur ini telah dilakukan memberikan hasil yang menjanjikan dan berguna untuk mengoptimalkan *Decision Tree*. Atribut tambahan adalah IPK. Secara keseluruhan, atribut berikut digunakan setelah satu: jenis kelamin, status pekerjaan, IPS_1, IPS_2, IPS_3, IPS_4, IPK, usia dan perpindahan. Bandingkan tanpa status pekerjaan.

2. METODOLOGI PENELITIAN



Gambar 1. Tahapan Penelitian

Gambar 1 memperlihatkan tahapan-tahapan dalam penelitian. Berikut adalah penjelasan mengenai setiap tahapannya:

- Identifikasi Masalah:** Tahap awal penelitian di mana peneliti mengidentifikasi permasalahan dalam bidang data mining, khususnya terkait model prediksi data. Tujuannya adalah memahami berbagai metode yang relevan serta memperoleh gambaran mengenai model klasifikasi dan prediksi.
- Rumusan Masalah:** Berdasarkan identifikasi masalah, rumusan penelitian ini adalah menentukan model yang paling efektif untuk memprediksi kelulusan mahasiswa tepat waktu serta mengimplementasikannya menggunakan algoritma Decision Tree.
- Tinjauan Studi:** Tahap ini mencakup kajian literatur mengenai prediksi kelulusan tepat waktu, algoritma Decision Tree, dan metodologi CRISP-DM. Selain itu, dilakukan analisis terhadap penelitian terdahulu, baik dalam bentuk jurnal nasional, internasional, maupun tesis yang berkaitan dengan data mining.
- Metode Penelitian:** Peneliti menentukan jenis metode penelitian yang akan digunakan dalam studi ini.
- Metode Pemilihan Sampel:** Tahap ini meliputi penentuan teknik pengambilan sampel yang sesuai untuk penelitian.
- Instrumentasi:** Peneliti menyiapkan alat atau instrumen yang diperlukan untuk pengumpulan data.

- g. Pemodelan: Dilakukan pemilihan algoritma yang paling sesuai dan memiliki tingkat akurasi tertinggi berdasarkan analisis sebelumnya.
- h. Pembuatan Prototipe: Berdasarkan hasil kajian dan analisis penelitian terdahulu, pada tahap ini dikembangkan model prediksi kelulusan mahasiswa tepat waktu.
- i. Pengujian Model Data: Proses pengujian diawali dengan pengumpulan data mahasiswa yang telah lulus dari *database* universitas di Jakarta. Selanjutnya, dilakukan klasifikasi dan pengukuran efektivitas prediksi menggunakan metode akurasi, *precision*, *recall*, dan *f1 score*.
- j. Pembuatan Laporan: Tahap akhir penelitian di mana peneliti menyusun laporan hasil penelitian serta menyimpulkan temuan dari evaluasi dan pengujian yang telah dilakukan.

2.1 Teknik Analisis, Prototipe Model dan Pengujian

Data yang digunakan dalam penelitian ini terutama tentang siswa manajemen, salah satu universitas di Jakarta, yang mencakup periode dari 2019 hingga 2023. Data dipilih karena mereka memberikan informasi yang sangat lengkap dan relevan untuk analisis yang dilakukan dalam analisis ini. Studi ini menggunakan metode untuk mengakomodasi aliran diagram *CRISP-DM* yang jelas (proses standar industri lintas industri untuk penambangan data) [19], [20], [21].

Model *CRISP-DM* yang renyah diakui sebagai standar industri untuk proses penambangan data dan secara sistematis menyediakan kerangka kerja terstruktur untuk memahami dan memproses data. Dengan menggunakan model ini, para peneliti dapat memastikan bahwa setiap langkah proses penambangan data dilakukan secara akurat dan efisien. Prosedur Penelitian dan Penerapan Algoritma Struktur Pengambilan Keputusan Menurut model *CRISP-DM* yang jelas yang dijelaskan pada Gambar 2. Ilustrasi ini divisualisasikan dengan jelas oleh langkah-langkah yang dilakukan dalam penelitian ini dari pengumpulan data hingga analisis akhir. Panduan ini akan membantu Anda mengejar dan lebih memahami proses penelitian.



Gambar 2. *CRISP-DM FRAMEWORK*

2.2 Feature Selection

Tujuan pemilihan fitur adalah untuk mengidentifikasi dan memilih sub kumpulan fitur yang paling bermakna untuk model prediktif. Proses ini penting karena dapat meningkatkan performa model, mengurangi *overfitting*, mengurangi waktu pelatihan, dan memfasilitasi interpretasi hasil model. Salah satu metode pemilihan fitur yang paling populer adalah *SelectKBest*, metode pemilihan fitur berbasis statistik.

SelectKBest adalah metode untuk memilih fitur terbaik berdasarkan skor statistik. Metode ini menggunakan berbagai uji statistik untuk menilai hubungan antara setiap fitur dan variabel target. *SelectKBest* memilih fitur dengan skor tertinggi, mengabaikan fitur lainnya.

Penerapan *SelectKBest* sangat relevan dalam konteks analisis data yang melibatkan *dataset* dengan banyak fitur, di mana tidak semua fitur berkontribusi secara signifikan terhadap model. Misalnya, dalam pengolahan data kesehatan, fitur seperti usia, berat badan, dan riwayat medis mungkin memiliki pengaruh yang berbeda terhadap hasil diagnosa penyakit dibandingkan dengan fitur lainnya seperti jenis kelamin atau lokasi geografi. Dengan menggunakan *SelectKBest*, kita dapat memastikan bahwa hanya fitur-fitur yang paling relevan yang digunakan dalam model, sehingga meningkatkan akurasi dan efisiensi model.

$$F = \frac{\text{Varians antar kelompok}}{\text{Varians dalam kelompok}} \quad (1)$$

Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi dan menerapkan metode *SelectKBest* dalam konteks. Dengan memahami dan memanfaatkan kekuatan metode ini diperoleh hasil keakuratan pengaruh paling tinggi sehingga lebih baik dan lebih efisien dalam proses analisis data dan pengembangan model prediktif [5].

2.3 Decision Tree

Algoritma pohon keputusan adalah salah satu teknik pembelajaran mesin yang paling populer untuk masalah klasifikasi dan regresi. Ini adalah model prediktif yang menggunakan struktur pohon keputusan untuk membuat keputusan berdasarkan serangkaian aturan hierarkis. Setiap simpul pada pohon mewakili keputusan berdasarkan nilai

atribut, dan setiap cabang mewakili hasil keputusan ini. Ada beberapa metode seperti C4.5, CART, ID3, dan CHAID [22], [23], [24], [25], [26].

Proses pembuatan pohon keputusan dimulai menggunakan pemilihan atribut yg paling informatif buat sebagai *node* pemisah pertama. Pemilihan atribut ini umumnya didasarkan dalam metrik misalnya *Information Gain*, *Gini Impurity*, atau *Gain Ratio*, yang mengukur seberapa baik atribut tadi memisahkan data ke pada kelas yang berbeda.

$$Gini(P) = \sum_{k=1}^k P_k(1 - P_k) = 1 - \sum_{k=1}^k P_k^2 \quad (2)$$

Setelah atribut pemisah pertama dipilih, data dipisahkan dari dalam nilai-nilai atribut tadi. Setiap cabang pada pohon mewakili nilai eksklusif berdasarkan atribut pemisah pertama, & setiap cabang bisa mempunyai anak-anaknya sendiri yang mewakili pemisahan lebih lanjut dari dalam atribut lain. Proses ini dilakukan secara *rekursif* buat setiap cabang, menggunakan pemilihan atribut terbaik buat setiap *node* berikutnya, sampai syarat berhenti terpenuhi. Kondisi berhenti bisa berupa mencapai kedalaman maksimum pohon, mencapai jumlah sampel minimum pada sebuah *node*, atau saat seluruh sampel pada sebuah *node* termasuk pada satu kelas.

3. HASIL DAN PEMBAHASAN

Aplikasi desktop ini dikembangkan menggunakan bahasa pemrograman Python dengan memanfaatkan beberapa *library*, seperti *Tkinter* untuk antarmuka pengguna (*GUI*), serta *Pandas*, *Matplotlib*, dan *Scikit-learn* untuk pemrosesan data serta pelatihan model. Aplikasi ini memiliki beberapa menu utama, di antaranya Load Data, yang memungkinkan pengguna mengimpor *dataset* dalam format *CSV* atau *Excel*; *Preprocess* Data, yang digunakan untuk membersihkan dan mentransformasi data, termasuk menangani nilai yang hilang, normalisasi, serta *encoding* kategori; *Visualize* Data, yang menyediakan berbagai pilihan visualisasi seperti histogram, *scatter plot*, dan *heatmap* untuk analisis eksploratif; serta *Train Model*, yang memungkinkan pengguna melatih model *machine learning* dengan parameter yang dapat disesuaikan.

Aplikasi ini juga menyediakan menu *Train Model*, yang memfasilitasi pengguna dalam menguji kinerja model yang telah dilatih dengan menggunakan berbagai metrik evaluasi, seperti akurasi. Fitur ini memungkinkan pengguna untuk mengevaluasi keandalan model sebelum menggunakannya pada data baru. Dengan antarmuka yang *user-friendly* dan fitur yang lengkap, aplikasi ini dirancang untuk memudahkan pengguna, baik yang memiliki latar belakang teknis maupun non-teknis, dalam melakukan analisis data dan mengembangkan model *machine learning* secara efektif. Aplikasi ini dirancang untuk memfasilitasi pengguna dalam melakukan analisis data serta membangun model prediktif secara efektif, tanpa memerlukan keahlian teknis atau pemahaman mendalam mengenai pemrograman. Melalui antarmuka yang intuitif dan fitur-fitur yang terstruktur dengan baik, aplikasi ini memungkinkan individu dari berbagai latar belakang untuk mengolah, mengevaluasi, dan menginterpretasikan data secara sistematis. Dengan demikian, aplikasi ini bertujuan untuk mengurangi hambatan teknis dalam analisis data, sehingga memungkinkan pengambilan keputusan berbasis data dapat dilakukan secara lebih luas dan efisien.

3.1 Pengelompokkan Data

Langkah pertama yang perlu Anda lakukan untuk memprediksi penyelesaian gelar mahasiswa secara tepat waktu adalah dengan mengelompokkan data Anda. Pengelompokan data harus dikategorikan secara sistematis agar dapat dianalisis menggunakan teknik pembelajaran mesin. Dalam hal ini digunakan teknik klasifikasi pohon keputusan. Teknik ini penting karena memberikan gambaran yang jelas tentang variabel-variabel yang mempengaruhi penyelesaian siswa tepat waktu. Pohon keputusan memungkinkan Anda merepresentasikan data dalam bentuk pohon keputusan yang mudah dipahami sehingga memungkinkan Anda mengidentifikasi pola dan tren.

Data yang digunakan dalam penelitian ini diperoleh dari Pusat Komputer dan Inovasi Digital (PusKID) di salah satu universitas di Jakarta. Data tersebut mencakup informasi tentang mahasiswa yang lulus pada tahun 2019, 2020, 2021, 2022, dan 2023, serta berkaitan dengan sistem kelulusan tepat waktu. Secara keseluruhan, terdapat 308 data mentah yang masih perlu diproses sebelum dianalisis lebih lanjut. Adapun dataset dapat dilihat pada Gambar 3.

	A	B	C	D	E	F	G	H
1	jenis_kelamin	usia	status_kerja	ipk	ipk_tahun_lulus	mtsk_akhir	tgl_lahir	nama_msk
2	P	28	Hanya Kuliah	4, 3,5, 3,46, 3,94, 3,73, 4, 4, 4	3,79	2017	2014	2103096 (Dasar Akutansi 1-6), (Dasar-dasar Ekonomi Mikro-4), (Dasar Das
3	P	28	Hanya Kuliah	1,65, 0,27, 2,14, 1,24, 0,4, 2,8, 2,38, 1,22, 3, 2,37, 1,5, 3,68, 2,65, 3, 3,96	3,04	2021	2014	0605096 (Dasar Akutansi 1-6), (Dasar-dasar Ekonomi Mikro-4), (Dasar Das
4	P	28	Hanya Kuliah	3,4, 3,35, 3, 3,67, 3,74, 3,81, 4, 4	3,58	2017	2014	1400206 (Dasar Akutansi 1-3), (Dasar-dasar Ekonomi Mikro-4), (Dasar Das
5	P	28	Hanya Kuliah	3,4, 3,23, 3,29, 3,43, 3,71, 4, 4	3,61	2017	2014	0607096 (Dasar Akutansi 1-3), (Dasar-dasar Ekonomi Mikro-4), (Dasar Das
6	P	27	Hanya Kuliah	2,4, 2,11, 2,11, 2, 2,5, 2,45, 2,6, 3,06, 3,61, 3,65, 4	2,99	2019	2014	0301096 (Dasar Akutansi 1-2), (Dasar-dasar Ekonomi Mikro-2), (Dasar Das
7	P	30	Hanya Kuliah	3,7, 3,5, 3,75, 3,65, 3,62, 3,75, 3, 3	3,63	2017	2014	0607094 (Dasar Akutansi 1-4), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
8	P	30	Hanya Kuliah	3,65, 3,71, 3,54, 3,8, 3,95, 3,95, 3,45, 3	3,67	2017	2014	0204496 (Dasar Akutansi 1-4), (Dasar-dasar Ekonomi Mikro-4), (Dasar Das
9	L	30	Hanya Kuliah	2,25, 2,72, 2,14, 2,17, 2,41, 2,72, 2,28, 2,71, 1,5, 0, 3, 0, 2, 3,3, 3,3	2,75	2021	2014	0707094 (Dasar Akutansi 1-6), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
10	L	28	Hanya Kuliah	3,4, 2,71, 2,62, 3,11, 2,63, 3,17, 3,27, 4, 3	3,04	2018	2014	1104496 (Dasar Akutansi 1-4), (Dasar-dasar Ekonomi Mikro-2), (Dasar Das
11	P	27	Hanya Kuliah	3,15, 3,38, 3,04, 3,68, 3,65, 3,63, 3,64, 4	3,46	2017	2014	0112096 (Dasar Akutansi 1-3), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
12	P	28	Hanya Kuliah	2,7, 2,67, 2,81, 2,42, 2,2,36, 3,44, 2,79, 2,38, 3,7, 2,29, 2,3, 3,65, 3,32	3,04	2020	2014	2606096 (Dasar Akutansi 1-6), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
13	P	29	Hanya Kuliah	2,8, 2,71, 2,33, 3,11, 2,43, 2,22, 3,17, 2,36, 1,29, 4, 0, 3,3, 0, 0, 4	3,1	2021	2014	0510094 (Dasar Akutansi 1-2), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
14	P	28	Hanya Kuliah	3,45, 3,46, 2,95, 3,62, 3,55, 3,53, 4, 4	3,47	2017	2014	1612095 (Dasar Akutansi 1-4), (Dasar-dasar Ekonomi Mikro-4), (Dasar Das
15	P	28	Hanya Kuliah	3, 2,79, 2,67, 3, 3,67, 3,64, 4, 4	3,3	2017	2014	2104096 (Dasar Akutansi 1-4), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
16	L	28	Hanya Kuliah	3,25, 2,5, 0,86, 2,47, 1,67, 2,07, 3,44, 3,47, 3,28, 4, 4	3,15	2019	2014	1406096 (Dasar Akutansi 1-2), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
17	P	28	Hanya Kuliah	3,75, 3,19, 3,5, 3,54, 3,82, 4, 4, 4	3,66	2017	2014	0203096 (Dasar Akutansi 1-4), (Dasar-dasar Ekonomi Mikro-4), (Dasar Das
18	P	27	Hanya Kuliah	3,55, 3,38, 3, 3,3, 3,86, 4, 4, 0, 0, 0, 0, 0, 0, 3,7	3,52	2021	2014	0711096 (Dasar Akutansi 1-4), (Dasar-dasar Ekonomi Mikro-4), (Dasar Das
19	P	27	Hanya Kuliah	2,85, 1,81, 2,38, 1,82, 2,2, 2,11, 3,25, 2,38, 2,81, 2,41, 3,2	3,01	2019	2014	1801096 (Dasar Akutansi 1-3), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
20	P	27	Hanya Kuliah	3,3, 2,86, 2,85, 2,81, 2,86, 3,36, 3,36, 3, 3	3,2	2018	2014	1712096 (Dasar Akutansi 1-4), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
21	P	27	Hanya Kuliah	3,3, 2,86, 2,87, 3,11, 2,75, 3,33, 2,9, 2,3, 0, 0, 0, 0, 0, 3,3	3,19	2021	2014	2201096 (Dasar Akutansi 1-3), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
22	L	36	W	1,7, 2,27, 1,53, 1,79, 3, 2,47	2,69	2016	2014	0502088 (Dasar Akutansi 1-6), (Matematika Ekonomi-1), (Pendidikan Agam
23	L	28	Hanya Kuliah	3,15, 2,08, 2,44, 2,89, 2,73, 3,45, 4, 3, 3	3,39	2018	2014	0101096 (Dasar Akutansi 1-3), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
24	L	30	Hanya Kuliah	3, 2,7, 2,78, 2,62, 2,71, 3, 2,53, 3, 3, 0, 0, 0, 0, 3	3,02	2021	2014	1406094 (Dasar Akutansi 1-2), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
25	P	28	Hanya Kuliah	2,8, 3,1, 2,86, 2,9, 3,14, 3,18, 3,63, 0, 3	3,29	2018	2014	1705096 (Dasar Akutansi 1-6), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
26	P	30	Bekerja dan kuliah	2,65, 2,8, 2,86, 2,95, 2,67, 3,21, 3,53, 1,5, 4, 2,7, 3,3	3,03	2019	2014	1302094 (Dasar Akutansi 1-6), (Dasar-dasar Ekonomi Mikro-2), (Dasar Das
27	L	28	Hanya Kuliah	2,5, 2,29, 2,24, 1,94, 3,47, 2,18, 1,93, 3,67, 1,3, 2,15, 0, 4, 3,3	2,96	2021	2014	2505096 (Dasar Akutansi 1-6), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
28	L	28	Hanya Kuliah	2,55, 2,52, 1,76, 2,75, 1,86, 3,4, 2,46, 2,88, 3,5, 3,39	3,13	2018	2014	2212094 (Dasar Akutansi 1-6), (Dasar-dasar Ekonomi Mikro-3), (Dasar Das
29	P	27	Hanya Kuliah	3,4, 3,42, 3,5, 3,48, 3,88, 4, 4, 4	3,66	2017	2014	0509096 (Dasar Akutansi 1-4), (Dasar-dasar Ekonomi Mikro-2), (Dasar Das
30	L	28	Hanya Kuliah	2,4, 2,61, 1,86, 2,27, 1,94, 1,57, 2,31, 2,73, 2,87, 3,2, 1, 3, 3	2,67	2021	2014	2012095 (Dasar Akutansi 1-6), (Dasar-dasar Ekonomi Mikro-2), (Dasar Das

Gambar 3. Dataset

Langkah pemrosesan data dalam penelitian penambangan data melibatkan beberapa fase penting yang perlu dilakukan. Fase ini bertujuan untuk memastikan bahwa data yang digunakan dalam penelitian memiliki kualitas yang memadai. Tujuan lainnya adalah untuk membuat analisis yang akurat saat mengelompokkan data. Langkah-langkah terperinci adalah sebagai berikut:

- Ubah semua data ke dalam bentuk numerik untuk memudahkan analisis dan pemrosesan lebih lanjut.
- Membagi data setiap semester dari semester 1 sampai semester 8 menjadi atribut-atribut tersendiri untuk mendapatkan informasi yang lebih detail mengenai kemajuan akademik mahasiswa per semester.
- Ganti data *null* dengan nilai 0 agar data tersebut dapat dimasukkan dalam proses pemilihan fitur tanpa menimbulkan kesalahan dalam perhitungan.
- Hapus atribut data yang tidak relevan atau tidak perlu untuk analisis Anda, sehingga membuat proses pemrosesan data Anda lebih efisien dan hanya berfokus pada informasi penting.

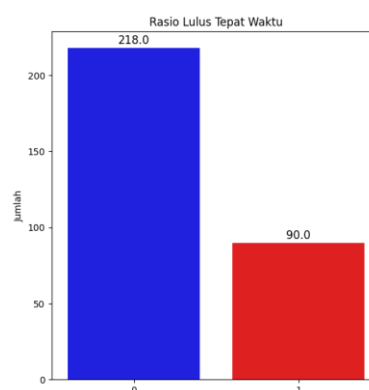
Jumlah data yang digunakan dalam penelitian ini adalah 308 data. Pertukaran data berlangsung dengan perbandingan 70:30. Alasan pertukaran data adalah: Pembagian 70% data digunakan sebagai data pelatihan. Data pelatihan ini digunakan untuk melatih algoritma pohon keputusan sehingga dapat mempelajari pola yang ada dan membangun model prediktif. 30% sisanya digunakan sebagai data uji. Data pengujian digunakan untuk mengevaluasi keakuratan model dalam mengklasifikasikan derajat dari waktu ke waktu berdasarkan data yang tidak dilatih sebelumnya.

Dataset yang digunakan dalam penelitian ini memiliki berbagai atribut, terdiri dari 15 atribut *prediktor* dan 1 atribut hasil. Atribut-atribut tersebut diperoleh dari sumber yang berhubungan dengan prediksi kelulusan mahasiswa tepat waktu, melalui wawancara serta analisis variabel yang diperkirakan berpengaruh dalam proses prediksi tersebut.

Proses pelatihan model membutuhkan sampel data yang dapat diolah untuk menghasilkan aturan (*rule*) klasifikasi. Beberapa atribut seperti jenis kelamin, usia, tahun ajaran, dan status kelulusan tepat waktu dikonversi ke dalam bentuk numerik. Sementara itu, atribut tahun ajaran, IPS, dan SKS dipecah berdasarkan semester, sedangkan mata kuliah dipisahkan berdasarkan masing-masing mata pelajaran.

3.2 Analisis Data

Selama fase analisis data, data yang digunakan dalam penelitian ini mencakup waktu yang tepat untuk gelar siswa. Berdasarkan data yang diterima, 218 siswa yang tidak menghabiskan waktu mereka dan 90 siswa berhasil diselesaikan. Persentase siswa yang selesai tepat waktu daripada siswa yang selesai tepat waktu dengan rasio sekitar 2,4:1 dan gambaran pertama waktu antara siswa, subjek dari objek penelitian dan penelitian, dan subjek objek penelitian; Objek penelitian membentuk dasar untuk analisis faktor - faktor yang dapat mempengaruhi keterlambatan dalam kesimpulan. Ini tertunda di akhir aspek akademik dan non-akademik. Temuan ini juga merupakan titik awal untuk pengembangan model prediksi temporal, dan dianalisis lebih lanjut dalam penelitian ini terlihat pada Gambar 4.



Gambar 4 Analisis Data

3.3. Feature Selection dengan SelectKBest

Dalam penelitian ini, kami melakukan langkah pemilihan fitur untuk memilih atribut yang paling relevan dengan variabel target, yaitu 'derajat'. Tujuan dari proses ini adalah untuk meningkatkan kinerja model pohon keputusan dengan menghilangkan fitur-fitur yang kurang penting dan mengurangi kompleksitas model. *Dataset* kelulusan mahasiswa memiliki 74 karakteristik. Pemilihan fitur dilakukan menggunakan *SelectKBest* dan 10 fitur dipilih. Di bawah ini adalah hasil untuk 10 fitur yang dipilih. Fitur yang dipilih adalah SKS_8, ips_3, ips_2, Pendidikan Pancasila dan Kewarganegaraan, SKS_1, Teknologi Informasi dan Komunikasi Terapan, ips_4, Bahasa Arab, Dasar-dasar Manajemen, dan Percakapan Bahasa Inggris. Fungsi yang tidak dipilih adalah ips_4, ips_2, ips_3, ips_5, ips_6, ips_7, ips_1, ips_3, ips_5, ips_6, ips_7, ips_8. Mata kuliah selain IPK, jenis kelamin, umur, kelas_1, kelas_2, kelas_3, kelas_4, kelas_5, kelas_6, tahun ajaran_7, kelas_8, Pendidikan Pancasila dan kewarganegaraan, Teknologi Informasi dan Komunikasi Terapan, Bahasa Arab, Dasar-dasar Manajemen, Bahasa Inggris Percakapan, Lampiran Lihat 3. IPK

Anda tidak berpengaruh karena skor fitur terbaik Anda tidak lebih tinggi dari skor percakapan bahasa Inggris Anda. Oleh karena itu, IPK tidak dijadikan sebagai variabel *prediktor*.

3.4. Klasifikasi Dengan Algoritma CART

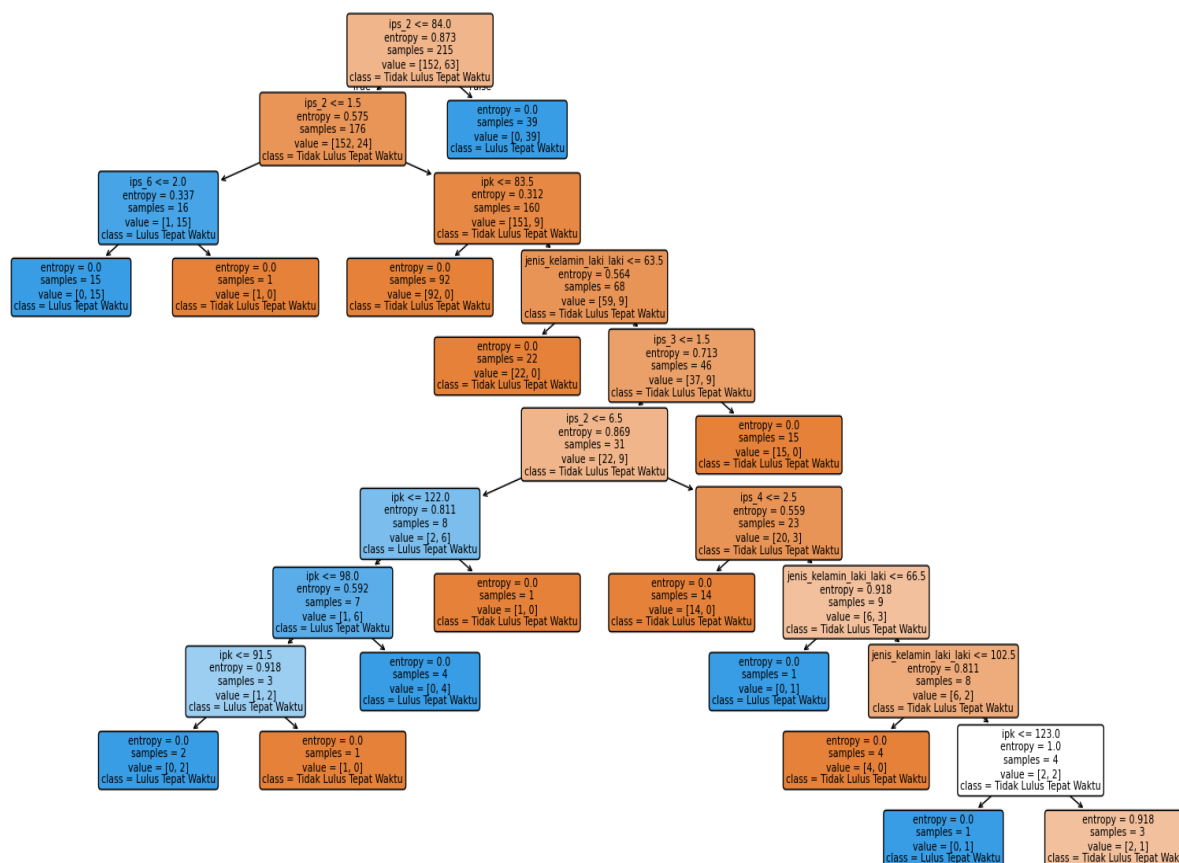
Model algoritma *CART* dikembangkan dengan menggunakan 70% dari data latih, yang diambil dari total 308 data yang telah dijelaskan sebelumnya. Proses pembentukan model dimulai dengan menentukan atribut akar dari data yang tersedia. Setelah itu, data diolah untuk menghitung nilai *Gini Index*.

Gini Index berperan penting dalam menilai kualitas informasi yang diperoleh dari atribut yang digunakan dalam proses klasifikasi. Setelah perhitungan *Gini Index* selesai, langkah berikutnya adalah menghitung nilai *Gini* untuk setiap atribut. Tujuan dari perhitungan ini adalah untuk mengukur sejauh mana ketidakpastian atau *impurity* dalam data berkurang ketika dilakukan pemisahan berdasarkan atribut yang dipilih.

Algoritma *CART* membangun model dengan membagi data ke dalam cabang-cabang berdasarkan atribut-atribut tertentu, sehingga terbentuk aturan-aturan sederhana. Dalam penelitian ini, model memanfaatkan nilai IPS semester sebagai kriteria utama dalam membedakan mahasiswa yang berpotensi lulus tepat waktu atau tidak. Penerapan algoritma dilakukan dengan langkah-langkah berikut:

- Menentukan fitur dengan *Gini Index* terbaik sebagai pembagi pertama.
- Membagi data ke dalam cabang-cabang sesuai nilai atribut.
- Mengulang proses pembagian hingga mencapai kondisi terminal (daun pohon) atau hingga *impurity* minimal.
- Menghasilkan aturan-aturan klasifikasi yang dapat digunakan untuk memprediksi data baru secara efisien.

Hasil dari proses klasifikasi akan menampilkan rekomendasi berupa struktur bangun pohon (*decision tree*) yang memberikan kesimpulan atau aturan-aturan prediktif. Dengan adanya bangun pohon ini, proses pengambilan keputusan terhadap data baru yang diinputkan dapat dilakukan lebih cepat dan efisien. Struktur pohon tersebut memvisualisasikan bagaimana setiap fitur atau variabel berkontribusi dalam menghasilkan prediksi, sehingga memudahkan pengguna untuk memahami alur logika yang digunakan oleh model. Bangun pohon yang dihasilkan dapat dilihat secara detail pada Gambar 3, yang menggambarkan hierarki keputusan mulai dari akar (*root*) hingga daun (*leaf*) yang merepresentasikan hasil akhir klasifikasi. Visualisasi ini tidak hanya membantu dalam interpretasi model tetapi juga memastikan transparansi dalam proses prediksi. Bangun pohon dapat terlihat pada Gambar 5



Gambar 5 Bangun Pohon

Dari bangun pohon tersebut dapat ditentukan kesimpulan yang aman lulus tepat waktu adalah:

- IPS semester 6 tidak kurang dari sama dengan 2,0
- IPS semester 3 tidak kecil sama dengan 1,5

Decision Tree merupakan salah satu algoritma *machine learning* yang beroperasi dengan cara membagi data menjadi beberapa subset berdasarkan nilai fitur tertentu. Dalam konteks prediksi kelulusan, *Decision Tree* mampu menghasilkan pohon keputusan yang menggambarkan aturan-aturan seperti, "Jika IPS semester 6 $\geq 2,0$ dan IPS semester 3 $\geq 1,5$, maka mahasiswa tersebut memiliki peluang tinggi untuk lulus tepat waktu."

Decision Tree akan mengembangkan model prediksi berdasarkan aturan-aturan yang terbentuk. Sebagai contoh, dari pohon keputusan yang dibuat, terungkap bahwa mahasiswa dengan IPS semester 6 $\geq 2,0$ dan IPS semester 3 $\geq 1,5$ memiliki peluang tinggi untuk lulus tepat waktu. Aturan ini dapat dijadikan acuan baik bagi mahasiswa maupun institusi pendidikan dalam menerapkan tindakan pencegahan, seperti memberikan dukungan akademik tambahan kepada mahasiswa yang belum memenuhi persyaratan tersebut.

3.5 Komparasi *Feature Selection* dengan murni *Decision Tree*

Hasil perbandingan yang diperoleh menunjukkan adanya perbedaan signifikan dalam tingkat akurasi prediksi. Model yang menerapkan *feature selection* mencapai akurasi, *precision*, *recall*, dan *f1 score* 97%, lebih tinggi dibandingkan dengan model tanpa *feature selection*, yang hanya memperoleh akurasi, *precision*, *recall*, dan *f1 score* 93%.

Temuan ini menunjukkan bahwa penggunaan *feature selection* dapat secara signifikan meningkatkan kinerja model dengan memilih fitur yang lebih relevan serta mengeliminasi *noise* dari data yang kurang penting. Dengan demikian, penerapan *feature selection* terbukti berperan penting dalam meningkatkan akurasi prediksi kelulusan mahasiswa tepat waktu, sehingga mendukung efektivitas model dalam menghasilkan keputusan yang lebih akurat.

Perbedaan tingkat akurasi antara model yang menggunakan *feature selection* dan model murni *Decision Tree* menunjukkan bahwa proses seleksi fitur berhasil memilah informasi yang paling berdampak pada target prediksi. Dalam studi ini, fitur seperti IPS semester 6 dan IPS semester 3 terbukti memiliki hubungan yang kuat dengan kelulusan tepat waktu. Dengan menyingkirkan fitur-fitur yang kurang relevan, model menjadi lebih terfokus pada variabel-variabel kunci, sehingga menurunkan risiko *overfitting* dan meningkatkan kemampuan generalisasi model. Selain itu, proses komputasi juga menjadi lebih ringkas karena jumlah fitur yang digunakan lebih sedikit, sementara kualitas prediksi tetap terjaga atau bahkan meningkat.

Temuan ini membawa implikasi signifikan bagi institusi pendidikan dalam menyusun strategi untuk mendukung kelulusan mahasiswa. Dengan memanfaatkan model yang telah melalui proses *feature selection*, institusi dapat lebih akurat mengidentifikasi faktor-faktor kritis yang memengaruhi kelulusan tepat waktu. Sebagai contoh, jika IPS semester 6 dan semester 3 terbukti menjadi faktor penentu utama, universitas dapat merancang program bimbingan akademik yang lebih spesifik untuk membantu mahasiswa meningkatkan performa mereka pada semester-semester tersebut. Selain itu, model ini juga dapat berfungsi sebagai alat pemantauan dini untuk mendeteksi mahasiswa yang berpotensi tidak lulus tepat waktu, sehingga tindakan intervensi dapat dilakukan secepat mungkin.

4. KESIMPULAN

Berdasarkan permasalahan yang telah diidentifikasi, survei literatur, tinjauan subjek penelitian, kerangka pemikiran, dan metodologi yang digunakan dalam penelitian prediksi kelulusan tepat waktu mahasiswa dengan teknik pohon keputusan dan optimasi pemilihan fitur, dapat disimpulkan bahwa model prediksi yang dikembangkan menggunakan *dataset* 308 siswa (218 gagal lulus tepat waktu dan 90 lulus sesuai jadwal) menghasilkan akurasi 97,85% dengan seleksi fitur (*SelectKBest*) dan 93,54% tanpa seleksi fitur, menunjukkan peningkatan signifikan sebesar 4,31%, di mana evaluasi model membuktikan bahwa pemilihan fitur tidak hanya meningkatkan akurasi tetapi juga *precision*, *recall*, dan *f1-score* dengan signifikan sehingga menghasilkan model yang lebih optimal melalui seleksi 10 atribut paling relevan yang berkontribusi besar dalam prediksi kelulusan tepat waktu, memberikan implikasi penting bagi pengembangan model prediksi di masa depan dengan rekomendasi implementasi langsung di universitas guna meningkatkan akurasi identifikasi mahasiswa berisiko dan intervensi tepat waktu melalui pemantauan berkelanjutan berbasis fitur terpilih agar model tetap relevan dan efektif secara *real-time*.

REFERENCES

- [1] W. L. W. Foh, S. L. Ang, C. Y. Lim, A. A.-L. Alaga, and G. H. Yeap, "Prediction of Tuberculosis Patients' Treatment Outcomes Using Multinomial Naive Bayes Algorithm and Class-Imbalanced Data," in *2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET)*, 2023, pp. 1–6. doi: 10.1109/GlobConET56651.2023.10150132.
- [2] Hartatik, K. Kusriani, and A. Budi Prasetyo, "Prediction of Student Graduation with Naive Bayes Algorithm," in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, IEEE, Nov. 2020, pp. 1–5. doi: 10.1109/ICIC50835.2020.9288625.
- [3] I. Buchori, "PERATURAN BADAN AKREDITASI NASIONAL PERGURUAN TINGGI NOMOR 22 TAHUN 2022," Jakarta, May 2022.
- [4] W. Agwil, H. Fransiska, and N. Hidayati, "Analisis Ketepatan Waktu Lulus Mahasiswa Dengan Menggunakan Bagging Cart," *FIBONACCI: Jurnal Pendidikan Matematika dan Matematika*, vol. 6, no. 2, p. 155, Dec. 2020, doi: 10.24853/fbc.6.2.155-166.
- [5] M. F. Thoriq, W. J. Pranoto, and F. Faldi, "Penerapan Seleksi Fitur Analysis of Variance Pada Algoritma Random Forest Classifier Dalam Klasifikasi Nilai Mahasiswa," *Explore: Jurnal Sistem Informasi dan Telematika*, vol. 14, no. 2, p. 185, Dec. 2023, doi: 10.36448/jsit.v14i2.3187.

- [6] E. Alpaydin, *Introduction to Machine Learning*, Fourth edition. The MIT Press, 2020.
- [7] A. Ermillian and K. Nugroho, “Perancangan Model Deteksi Potensi Siswa Putus Sekolah Menggunakan Metode Logistic Regression Dan Decision Tree,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 9, no. 3, pp. 281–295, Dec. 2024, doi: 10.30591/jpit.v9i3.8007.
- [8] F. Balouchzahi, H. Lakshmaiah Shashirekha, and G. Sidorov, “Ensembled Feature Selection for Urdu Fake News Detection,” 2021. [Online]. Available: <https://mangaloreuniversity.ac.in/dr-h-l-shashirekha>
- [9] R. Wibowo and H. Indriyawati, “Top-k Feature Slection Untuk Deteksi Penyakit Hepatitis Menggunakan Algoritme Naïve Bayes,” *Jurnal Buana Informatika*, vol. 11, no. 1, pp. 1–9, May 2020, doi: 10.24002/jbi.v11i1.2456.
- [10] Rovidatul, Y. Yunus, and G. W. Nurcahyo, “Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 193–199, Apr. 2023, doi: 10.37859/coscitech.v4i1.4755.
- [11] P. Patlisan, “Optimasi Akurasi Model Decision Tree Menggunakan Random Forest Regression untuk Prediksi Kuantitas Pembelian Barang pada Perusahaan Manufaktur,” *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 14, no. 2, pp. 217–228, Nov. 2023, doi: 10.24176/simet.v14i2.9494.
- [12] M. Yağcı, “Educational data mining: prediction of students’ academic performance using machine learning algorithms,” *Smart Learning Environments*, vol. 9, no. 1, p. 11, 2022, doi: 10.1186/s40561-022-00192-z.
- [13] Y. Jang, S. Choi, H. Jung, and H. Kim, “Practical early prediction of students’ performance using machine learning and eXplainable AI,” *Educ Inf Technol (Dordr)*, vol. 27, pp. 1–35, Jun. 2022, doi: 10.1007/s10639-022-11120-6.
- [14] B. Mounika and V. Persis, “A Comparative Study of Machine Learning Algorithms for Student Academic Performance,” *International Journal of Computer Sciences and Engineering*, vol. 7, pp. 721–725, Apr. 2019, doi: 10.26438/ijcse/v7i4.721725.
- [15] M. Gunawan, M. Zarlis, and R. Roslina, “Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 513, Apr. 2021, doi: 10.30865/mib.v5i2.2925.
- [16] F. F. Anwar, A. I. Jaya, and M. Abu, “Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dengan Penerapan Algoritma C4.5,” *JURNAL ILMIAH MATEMATIKA DAN TERAPAN*, vol. 19, no. 1, pp. 19–28, Jun. 2022, doi: 10.22487/2540766x.2022.v19.i1.15880.
- [17] I. R. Lutunani and A. Nugroho, “Analisis Prediksi Mahasiswa Terhadap Kelulusan Tepat Waktu Menggunakan Metode Data Mining Decision Tree (Studi Kasus: FTI UKSW),” *Jurnal Teknologi Informasi dan Komunikasi*, vol. 7, no. 2, p. 2023, 2023, doi: 10.35870/jti.
- [18] C. Wirawan, “Teknik Data Mining Menggunakan Algoritma Decision Tree C4.5 untuk Memprediksi Tingkat Kelulusan Tepat Waktu,” *Applied Information Systems and Management (AISM)*, vol. 3, pp. 47–52, 2020.
- [19] P. L. C. P. Lopez, *DATA MINING. The CRISP-DM METHODOLOGY. The CLEM Language and IBM SPSS MODELER*. Lulu, 2021.
- [20] M. Dursun and M. Sabyrzhan, “CRISP - DM Process in Smart Building Management,” in *2024 11th International Conference on Electrical and Electronics Engineering (ICEEE)*, 2024, pp. 310–315. doi: 10.1109/ICEEE62185.2024.10779257.
- [21] M. T. H. Suhendar and Y. Widayani, “Machine Learning Application Development Guidelines Using CRISP-DM and Scrum Concept,” in *2023 IEEE International Conference on Data and Software Engineering (ICoDSE)*, 2023, pp. 168–173. doi: 10.1109/ICoDSE59534.2023.10291438.
- [22] M. Zha, L. Zhu, Y. Zhu, and T. Hu, “Link Quality Analysis of Wireless Sensor Networks Based on CART Decision Tree,” in *2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, 2023, pp. 600–605. doi: 10.1109/ICICML60161.2023.10424860.
- [23] R. Tang and X. Zhang, “CART Decision Tree Combined with Boruta Feature Selection for Medical Data Classification,” in *2020 5th IEEE International Conference on Big Data Analytics (ICBDA)*, 2020, pp. 80–84. doi: 10.1109/ICBDA49040.2020.9101199.
- [24] C. J. D. Albia and D. Origines, “Selecting of Pet Adopters using C4.5 Decision Tree Model Algorithm,” in *2022 2nd International Conference in Information and Computing Research (iCORE)*, 2022, pp. 30–35. doi: 10.1109/iCORE58172.2022.00025.
- [25] M. Yin, “Application of Decision Tree Classification ID3 Algorithm in Learning Strategy Recommendation of Online Teaching Platform,” in *2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, 2023, pp. 1–4. doi: 10.1109/SMARTGENCON60755.2023.10442004.
- [26] M. Tian, “Construction cost optimization technology based on decision tree CHAID algorithm,” in *The International Conference on Forthcoming Networks and Sustainability (FoNeS 2022)*, 2022, pp. 583–587. doi: 10.1049/icp.2022.2515.