

SMARTGRAD: Prediksi Kelulusan Tepat Waktu Mahasiswa Kampus Merdeka

Agung Wibowo^{1*}, Ade Pratama², Dwi Setiawan³

¹²Teknik Informatika, Fakultas Komputer dan Pendidikan, Universitas Ngudi Waluyo, Ungaran, Indonesia

³Teknologi Permainan, ³Sekolah Tinggi Multi Media Yogyakarta, Indonesia

^{1*}agungwibowo@unw.ac.id, ²adepratama@unw.ac.id, ³dwis0609@gmail.com

ABSTRACT – On-time graduation is a primary indicator of student success and serves as a key benchmark for the quality of higher education institutions. This study aims to develop SmartGrad, a prediction model for on-time graduation based on the Naive Bayes algorithm, supported by feature selection using Decision Tree. The model integrates academic variables (semester GPA, average grades) and non-academic variables (types of MBKM, employment status, age) to produce accurate and contextual predictions. The research dataset comprises 313 entries with 17 attributes, processed through feature selection and classification stages. Evaluation results demonstrate the model's excellent performance, with an average accuracy of 88.8%, precision of 90.5%, recall of 97.9%, and an F1-score of 94.0%. The implementation of SmartGrad as an interactive web application based on Streamlit supports transparent and easily comprehensible decision-making. The novelty of this research lies in the integration of MBKM factors and employment status into the prediction model, as well as the application of an interpretable AI approach to support higher education policies and the achievement of Sustainable Development Goal 4 (Quality Education). These findings are expected to serve as a strategic reference for higher education administrators in enhancing academic quality and the effectiveness of the Freedom of Learning Independent Campus program.

Keywords: Decision Tree; Graduation; Kampus Merdeka; Naïve Bayes; Streamlit.

ABSTRAK – Kelulusan tepat waktu merupakan indikator utama keberhasilan studi mahasiswa dan menjadi tolok ukur mutu perguruan tinggi. Penelitian ini mengembangkan SmartGrad, sebuah model prediksi kelulusan tepat waktu berbasis algoritma Naive Bayes dengan dukungan seleksi fitur menggunakan Decision Tree. Model ini mengintegrasikan variabel akademik (IP semester, rata-rata nilai) dan non-akademik (jenis MBKM, status pekerjaan, usia) untuk menghasilkan prediksi yang akurat dan kontekstual. Dataset penelitian terdiri dari 313 entri dengan 17 atribut yang diproses melalui tahapan seleksi fitur dan klasifikasi. Hasil evaluasi menunjukkan performa model yang sangat baik dengan akurasi rata-rata 88,8%, precision 90,5%, recall 97,9%, dan F1-score 94,0%. Implementasi SmartGrad dalam bentuk aplikasi web interaktif berbasis Streamlit mendukung pengambilan keputusan yang transparan dan mudah dipahami. Kebaruan penelitian ini terletak pada integrasi faktor Merdeka Belajar Kampus Merdeka (MBKM) dan status pekerjaan ke dalam model prediksi serta penerapan pendekatan interpretable AI untuk mendukung kebijakan pendidikan tinggi dan pencapaian SDGs Tujuan 4 (Pendidikan Berkualitas). Temuan ini diharapkan menjadi referensi strategis bagi pengelola perguruan tinggi dalam meningkatkan mutu akademik dan efektivitas program Kampus Merdeka.

Kata Kunci: Decision Tree; Kelulusan; Kampus Merdeka; Naïve Bayes; Streamlit.

1. PENDAHULUAN

Kelulusan tepat waktu merupakan indikator penting dalam keberhasilan studi mahasiswa. Selain memberikan manfaat langsung bagi mahasiswa seperti efisiensi biaya dan percepatan karier, hal ini juga berdampak positif terhadap reputasi dan akreditasi institusi pendidikan. Salah satu faktor utama yang memengaruhi kelulusan adalah capaian akademik, terutama Indeks Prestasi (IP) tiap semester. IP yang rendah berpotensi menyebabkan keterlambatan kelulusan.

Penelitian prediksi kelulusan mahasiswa sebelumnya [1] yang memanfaatkan algoritma Naïve Bayes dan Decision Tree (C4.5) untuk memprediksi ketepatan waktu kelulusan. Hasil penelitian tersebut menunjukkan bahwa kedua algoritma ini efektif digunakan dalam mendukung pengambilan

keputusan akademik berbasis data. Seperti dijelaskan oleh Ifenthaler dan Gibson dalam *Adoption of Data Analytics in Higher Education Learning and Teaching* [2], Penerapan analitik data di pendidikan tinggi mendukung pengambilan keputusan berbasis bukti untuk meningkatkan keberhasilan mahasiswa.

Sejalan dengan Tujuan 4 SDGs, pendidikan tinggi yang berkualitas dan inklusif menjadi prioritas global. Pemerintah Indonesia mendukung hal ini melalui program Merdeka Belajar Kampus Merdeka (MBKM), yang memberikan fleksibilitas pembelajaran dan pengalaman praktis kepada mahasiswa. Sejak diluncurkan pada tahun 2020, program MBKM telah diikuti oleh lebih dari 760.000 mahasiswa. Khusus untuk skema MBKM Mandiri, jumlah peserta meningkat dari 241.000 (2020–2022)



menjadi 257.000 pada tahun 2023, menunjukkan antusiasme yang tinggi terhadap program ini [3].

MBKM terdiri dari dua skema utama: (1) MBKM Flagship: Program unggulan dari Kemendikbudristek seperti MSIB, Kampus Mengajar, PMM, IISMA, Wirausaha Merdeka, dan Praktisi Mengajar. (2) MBKM Mandiri: Implementasi internal oleh perguruan tinggi, memungkinkan mahasiswa mengambil mata kuliah lintas prodi, proyek penelitian, pengabdian masyarakat, magang mandiri, aistensi mengajar. Kedua skema ini saling melengkapi dalam mendukung kelulusan tepat waktu dan peningkatan kualitas lulusan yang adaptif terhadap dunia kerja dan tantangan global. Data sebaran skema MBKM ini ditunjukkan pada tabel 1.1.

Tabel 1. Jumlah Mahasiswa Peserta MBKM[3]

TAHUN	MANDIRI	FLAGSHIP
<2022	241.000	177.000
2023	257.000	320.000
2024	500.000	193.000
2025	500.000	225.000

Dalam konteks Program Merdeka Belajar Kampus Merdeka (MBKM), fleksibilitas yang ditawarkan program ini memungkinkan mahasiswa untuk mengikuti berbagai kegiatan di luar kampus yang dapat memperkaya pengalaman belajar mereka. Namun, hal ini juga menambah tantangan dalam pengelolaan waktu dan kurikulum, yang dapat mempengaruhi kelulusan tepat waktu.

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma Naive Bayes dan Decision Tree banyak digunakan untuk memprediksi kelulusan mahasiswa, namun masing-masing memiliki keterbatasan. Penelitian pertama berkaitan dengan integrasi Project-Based Learning dengan Naive Bayes dan menemukan korelasi kuat terhadap prediksi kelulusan, tetapi tidak membandingkan algoritma lain dan detail data set kurang jelas[4]. Penelitian berikutnya membandingkan Naive Bayes dan KNN, dengan hasil akurasi 80% untuk Naive Bayes, namun variabel non-akademik belum dieksplorasi[5]. Penelitian lainnya menggunakan data set besar dan menguji beberapa algoritma, di mana metode ensemble unggul (84,03%), sedangkan Naive Bayes dan Decision Tree tetap bernilai untuk interpretabilitas, meskipun tidak dioptimalkan[6]. Generalisasi penelitian banyak membandingkan Decision Tree dan Naive Bayes dengan akurasi tinggi (hingga 93%) [7][8], tetapi sampel terbatas, validasi kurang jelas, dan variabel MBKM belum dimasukkan. Naive Bayes dan Decision Tree (C4.5) merupakan metode populer dalam prediksi kelulusan mahasiswa karena kemampuannya menangani data kategorial dan numerik serta interpretabilitas yang baik [1]. Naive Bayes dikenal dengan kesederhanaan dan efisiensi komputasi, sedangkan Decision Tree unggul dalam menghasilkan model yang mudah

dipahami [9][10]. Beberapa studi membandingkan kedua algoritma ini dan menemukan bahwa Decision Tree sering memberikan akurasi lebih tinggi dibanding Naive Bayes [11][12][13].

Pada sebagian besar penelitian masih terbatas pada variabel akademik seperti IPK, jumlah SKS, dan kehadiran, tanpa mempertimbangkan faktor non-akademik atau kebijakan pendidikan seperti Merdeka Belajar Kampus Merdeka (MBKM) [3]. Selain itu, keterbatasan ukuran sampel, kurangnya validasi yang kuat, dan minimnya analisis kontekstual membuat hasil penelitian sulit digeneralisasi [14]. Penelitian ini hadir untuk mengisi celah tersebut dengan mengintegrasikan variabel akademik dan non-akademik, termasuk pengalaman MBKM, status beasiswa, status kerja, dan rumpun ilmu, serta tetap menggunakan algoritma interpretable (Naive Bayes dan Decision Tree) untuk mendukung transparansi pengambilan keputusan. Dengan pendekatan ini, penelitian ini diharapkan tidak hanya meningkatkan akurasi prediksi, tetapi juga memberikan kontribusi strategis dalam mendukung kebijakan nasional dan pencapaian SDGs Tujuan 4 (Pendidikan Berkualitas) [2].

Kebaruan penelitian ini mengusulkan pendekatan baru dalam prediksi kelulusan tepat waktu mahasiswa dengan mengombinasikan algoritma Decision Tree dan Naive Bayes dalam satu kerangka kerja berbasis kecerdasan buatan (Artificial Intelligence). Decision Tree digunakan sebagai metode seleksi fitur untuk mengidentifikasi variabel-variabel yang paling relevan terhadap kelulusan mahasiswa, sedangkan Naive Bayes digunakan sebagai algoritma utama untuk melakukan prediksi berdasarkan fitur-fitur terpilih [11]. Pendekatan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa pemilihan atribut yang tepat dapat meningkatkan akurasi model prediksi [15]. Penelitian yang sudah ada menyampaikan bahwa performa kedua algoritma bergantung pada karakteristik dan data yang digunakan [16]. Model prediksi yang hanya mengandalkan data akademik seringkali gagal menangkap kompleksitas faktor eksternal yang memengaruhi perjalanan studi mahasiswa [17]. Berdasarkan beberapa evaluasi pada model sebelumnya, maka untuk model yang akan dikembangkan mengintegrasikan variabel akademik (IPK, jumlah SKS, semester) dan non-akademik (pengalaman MBKM, status beasiswa, status kerja, rumpun ilmu) untuk menghasilkan prediksi yang lebih akurat dan kontekstual. Integrasi variabel non-akademik ini penting karena kebijakan Merdeka Belajar Kampus Merdeka (MBKM) memberikan pengaruh signifikan terhadap pola studi mahasiswa [3][12]. Selain itu, faktor sosial-ekonomi dan pengalaman kerja juga terbukti memengaruhi ketepatan waktu kelulusan [7][8].

Model prediksi yang dikembangkan dalam penelitian ini memanfaatkan kombinasi variabel akademik, non-akademik, dan demografis untuk

memberikan hasil yang lebih akurat dan kontekstual. Variabel akademik mencakup Indeks Prestasi Kumulatif (IPK) pada semester 4, 5, dan 6 (PK4, IPK5, IPK6) yang memiliki nilai information gain tertinggi dalam pemodelan Decision Tree [1]. Selain itu, Indeks Prestasi Semester 5 (IPS5) digunakan untuk menangkap dinamika capaian akademik antar semester [1], sedangkan jumlah SKS yang telah ditempuh menjadi indikator penting untuk mengukur progres studi dan beban akademik [10].

Faktor non-akademik juga berperan signifikan dalam memengaruhi ketepatan waktu kelulusan. Partisipasi mahasiswa dalam program Merdeka Belajar Kampus Merdeka (MBKM), seperti magang, proyek desa, atau pertukaran pelajar, menjadi variabel penting karena dapat memengaruhi pola studi dan durasi penyelesaian kuliah [3]. Hal ini didukung oleh penelitian yang menunjukkan bahwa program pembelajaran berbasis pengalaman seperti magang dapat meningkatkan keterlibatan mahasiswa, namun jika tidak terstruktur dengan baik, berpotensi menunda kelulusan [18]. Status beasiswa turut diperhitungkan karena dukungan finansial sering dikaitkan dengan motivasi belajar yang lebih tinggi [7] dan memungkinkan mahasiswa untuk lebih fokus pada studi tanpa tekanan ekonomi [19]. Selain itu, status pekerjaan mahasiswa (bekerja atau tidak) dapat memengaruhi alokasi waktu belajar, sedangkan rumpun ilmu (misalnya teknik, bisnis, atau pendidikan) memengaruhi beban kurikulum dan pola kelulusan. Faktor demografis seperti jenis kelamin dan usia masuk kuliah juga relevan, karena keduanya dapat memengaruhi kedewasaan belajar dan kesiapan akademik [1]. Integrasi ketiga kelompok variabel ini diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai faktor-faktor yang memengaruhi kelulusan tepat waktu, sekaligus meningkatkan akurasi model prediksi. Implementasi sistem pembelajaran adaptif berbasis analytics telah berhasil meningkatkan retensi mahasiswa dan mengurangi risiko tidak lulus tepat waktu [20].

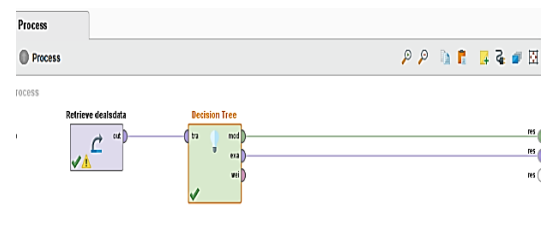
2. METODOLOGI

Penelitian ini menggunakan pendekatan kombinasi antara algoritma Decision Tree dan Naïve Bayes untuk membangun model prediksi kelulusan tepat waktu mahasiswa dalam program Kampus Merdeka. Dataset yang digunakan memiliki distribusi kelas yang tidak seimbang, yaitu 281 mahasiswa lulus tepat waktu (89,8%) dan 32 mahasiswa tidak lulus tepat waktu (10,2%). Ketidakseimbangan ini berpotensi menyebabkan bias terhadap kelas mayoritas. Untuk mengurangi dampak bias, penelitian ini menggunakan metrik evaluasi yang komprehensif (precision, recall, F1-score) dan memanfaatkan prior probabilities pada Naïve Bayes, serta validasi dengan cross-validation agar hasil lebih stabil. Metode ini terdiri dari dua tahapan utama,

yaitu seleksi fitur dan klasifikasi, yang dijelaskan sebagai berikut:

2.1. Seleksi Fitur dengan Decision Tree

Tahap pertama dalam proses analisis adalah seleksi fitur menggunakan algoritma Decision Tree. Tujuan dari tahap ini adalah untuk mengidentifikasi atribut-atribut yang paling relevan terhadap target prediksi, yaitu status kelulusan mahasiswa. Decision Tree bekerja dengan menghitung nilai Information Gain dari setiap atribut, yang menunjukkan seberapa besar kontribusi atribut tersebut dalam membagi data secara optimal. Tahapan seleksi fitur dengan menggunakan rapidminer ditunjukkan pada gambar 2.1.



Gambar 1. Model Decision Tree untuk Pemilihan Fitur Dengan Rapidminer

Rumus untuk menghitung nilai information gain;

$$IG(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \times Entropy(S_v) \quad (1)$$

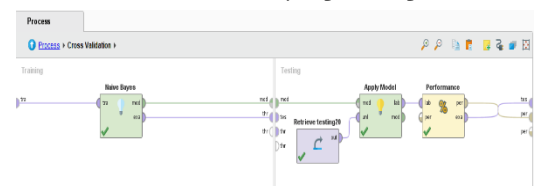
$$[Entropy(S) = - \sum_{i=1}^c p_i \log_2 p_i] \quad (2)$$

Attribute dengan nilai Information Gain tertinggi akan dipilih sebagai fitur utama dalam proses klasifikasi. Seleksi fitur ini bertujuan untuk menyederhanakan model, meningkatkan akurasi, dan mengurangi kompleksitas komputasi.

2.2. Prediksi Kelulusan dengan Naïve Bayes

Setelah fitur-fitur utama terpilih, tahap selanjutnya adalah klasifikasi menggunakan algoritma Naïve Bayes. Algoritma bekerja berdasarkan prinsip Teorema Bayes, dengan asumsi bahwa setiap fitur bersifat independen terhadap fitur lainnya. Design Naive bayes menggunakan cross validation pada rapidminer ditunjukkan pada gambar 2.2.

Gambar 2. Model Naive Bayes pada Rapidminer



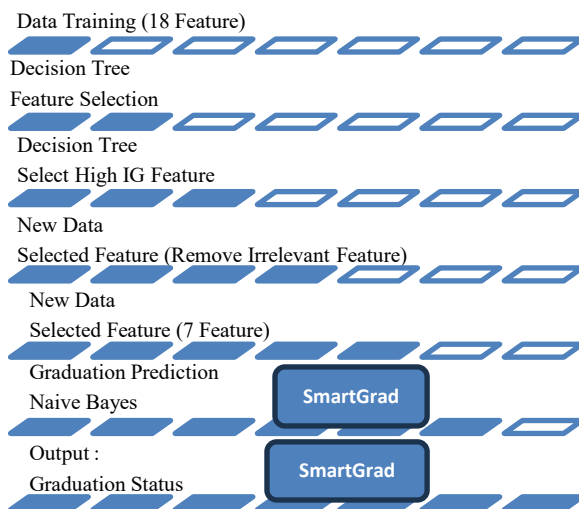
Model ini menggabungkan data training dari berbagai variabel akademik seperti IPS, Rerata nilai serta variabel non-akademik seperti partisipasi MBKM, status kerja, dan usia saat menjadi mahasiswa. Dengan pendekatan ini, model prediksi yang dihasilkan

diharapkan mampu memberikan hasil yang lebih akurat dan kontekstual dalam mendukung kebijakan pendidikan nasional, khususnya program Kampus Merdeka. Rumus dasar Naive Bayes adalah sebagai berikut :

$$[P(C_k | X) = \frac{P(C_k) \prod_{i=1}^n P(x_i | C_k)}{P(X)}] \quad (3)$$

Dalam konteks prediksi kelulusan mahasiswa, Naive Bayes mampu mengolah kombinasi data akademik dan non-akademik seperti IPK, jumlah SKS, status kerja, dan partisipasi MBKM untuk menghasilkan klasifikasi yang akurat [7][8]. Penelitian oleh terdahulu menunjukkan bahwa penerapan Naive Bayes dalam memprediksi kelulusan mahasiswa menghasilkan akurasi sebesar 94,31%, dengan precision 91% dan recall 95%, menjadikannya salah satu algoritma yang efektif dalam sistem pendukung keputusan akademik[9].

2.3. Tahapan Penelitian



2.4. Tahapan Penelitian

Langkah pertama adalah melakukan seleksi fitur menggunakan algoritma Decision Tree pada data training. Tujuannya adalah untuk mengevaluasi setiap atribut berdasarkan kontribusinya terhadap prediksi kelulusan mahasiswa. Decision Tree menghitung nilai Information Gain untuk menentukan seberapa besar suatu atribut mampu membagi data secara informatif.

Tabel 2 Fitur Data Training

No	Jenis Atribut	Keterangan
1	NIM	Nomor Induk Mahasiswa
2	Gender	Laki-laki atau Perempuan
3	USIA	Usia Masuk menjadi mahasiswa
4	SKS MBKM	Jumlah SKS MBKM
5	JUMLAH MATAKULIAH	Jumlah Matakuliah MBKM yang ditempuh
6	IP2	Indeks Prestasi Semester 2
7	IP3	Indeks Prestasi Semester 3
8	IP4	Indeks Prestasi Semester 4

9	IP5	Indeks Prestasi Semester 5
10	IP6	Indeks Prestasi Semester 6
11	BOBOT NILAI	Bobot nilai SKS x Nilai Huruf MBKM
12	RATA-RATA NILAI	Rata-Rata nilai bobot MBKM
13	PROGRAM MBKM	Magang, Asistensi Mengajar, MSIB
14	JENIS MBKM	Mandiri dan Flagship
15	BEASISWA	Mahasiswa Penerima Beasiswa
16	STATUS PEKERJAAN	Mahasiswa yang bekerja
17	RUMPUN ILMU	Rumpun ilmu mahasiswa peserta MBKM
18	SEMESTER TEMPUH	Semester yang ditempuh hingga selesai
19	KETEPATAN LULUS	Lulus Tepat waktu atau tidak tepat waktu

Setelah nilai Information Gain dihitung, dilakukan perangkingan dan hanya fitur-fitur dengan nilai tertinggi yang dipilih. Fitur-fitur ini dianggap paling relevan dan memiliki pengaruh signifikan terhadap hasil klasifikasi. Langkah ini bertujuan untuk menyederhanakan model dan meningkatkan akurasi. Hasil pemilihan fitur dengan rapidminer ditunjukkan pada gambar 2.4.

Tree

```

IP5 > 3.415
| IP3 > 2.775
| | IP2 > 3.805: TEPAT (TEPAT=70, TIDAK=0)
| | IP2 ≤ 3.805
| | | IP2 > 3.795: TEPAT (TEPAT=2, TIDAK=1)
| | | IP2 ≤ 3.795
| | | | mandiri/flagship = FLAGSHIP: TEPAT (TEPAT=21, TIDAK=0)
| | | | mandiri/flagship = MANDIRI
| | | | | rata-rata nilai > 15.400: TEPAT (TEPAT=15, TIDAK=0)
| | | | | rata-rata nilai ≤ 15.400
| | | | | BEKERJA/TIDAK = BEKERJA: TEPAT (TEPAT=60, TIDAK=2)
| | | | | BEKERJA/TIDAK = TIDAK
| | | | | USIAMASUK > 25.500
| | | | | | USIAMASUK > 41: TEPAT (TEPAT=5, TIDAK=0)
| | | | | | USIAMASUK ≤ 41: TIDAK (TEPAT=1, TIDAK=4)
| | | | | | USIAMASUK ≤ 25.500: TEPAT (TEPAT=72, TIDAK=11)
| | IP3 ≤ 2.775: TEPAT (TEPAT=1, TIDAK=1)
IP5 ≤ 3.415: TEPAT (TEPAT=35, TIDAK=12)
    
```

Gambar 3 Fitur yang dihasilkan dari rapidminer

Fitur-fitur yang memiliki nilai Information Gain rendah atau tidak memberikan kontribusi berarti terhadap prediksi akan dihapus. Penghapusan ini membantu mengurangi kompleksitas model, mempercepat proses komputasi, dan mencegah overfitting.

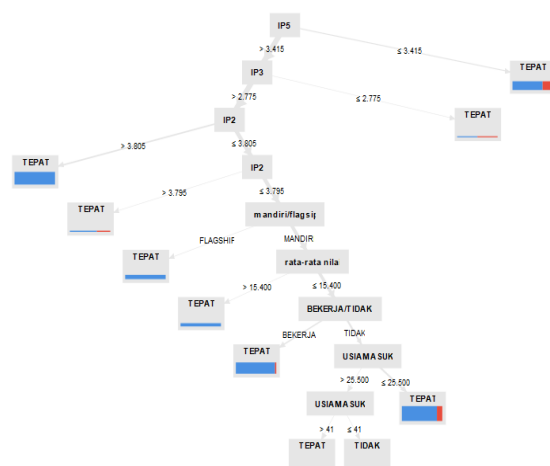
Setelah fitur-fitur utama ditentukan, data mahasiswa baru disiapkan hanya dengan menggunakan fitur-fitur yang telah diseleksi. Hal ini memastikan bahwa model prediksi bekerja secara konsisten dan fokus pada variabel yang paling berpengaruh. Data baru yang telah disederhanakan kemudian diproses menggunakan algoritma Naive Bayes. Algoritma ini menghitung probabilitas kelulusan mahasiswa berdasarkan fitur-fitur terpilih. Naive Bayes dipilih karena kesederhanaannya, efisiensi komputasi, dan kemampuannya menangani data kategorikal maupun numerik.

Selanjutnya dilakukan pengujian dengan menggunakan data testing yang belum diketahui classnya untuk mengetahui model yang dikembangkan dapat melakukan prediksi kelulusan dengan label lulus tepat waktu atau tidak.

Langkah terakhir adalah menghasilkan output berupa prediksi status kelulusan mahasiswa, apakah lulus tepat waktu atau tidak. Hasil ini dapat digunakan oleh institusi pendidikan untuk melakukan intervensi dini terhadap mahasiswa yang berisiko tidak lulus tepat waktu. Pada mengimplementasikan model prediksi ke dalam aplikasi web interaktif berbasis Streamlit. Tujuan implementasi ini adalah mempermudah pengelola perguruan tinggi dalam menggunakan model prediksi secara praktis. Fitur utama aplikasi meliputi: 1)Upload Data: Pengguna dapat mengunggah data mahasiswa baru. 2)Prediksi Kelulusan: Sistem menampilkan hasil prediksi apakah mahasiswa akan lulus tepat waktu atau tidak. 3)Visualisasi dan Interpretasi: Aplikasi menampilkan probabilitas prediksi dan fitur yang paling berpengaruh, mendukung prinsip interpretable AI.

3. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini sejumlah 313 data yang terbagi menjadi dua kategori yaitu sejumlah 281 kategori lulus tepat waktu dan 32 lulus tidak tepat waktu/mengalami keterlambatan. Jumlah data training dan testing menggunakan perbandingan 80:20. Data testing digunakan untuk mengetahui akurasi dari prediksi kelulusan. Fitur yang digunakan sejumlah 17 yang dibagi dari 16 sebagai fitur prediksi, dan 1 fitur sebagai label. Hasil percobaan decision tree menggunakan tools Rapidminer dapat dilihat pada gambar 3.1 berikut ini:

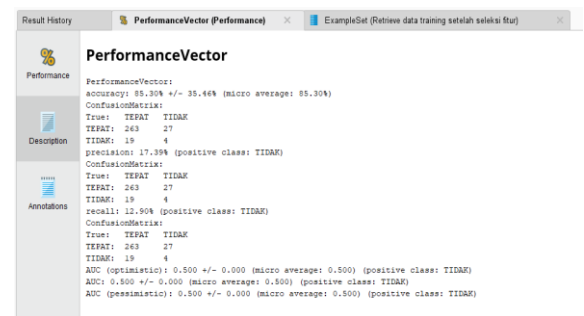


Gambar 4 Pohon Keputusan Seleksi Fitur menggunakan Rapidminer

Hasil perhitungan dengan rapidminer bahwa fitur yang memiliki nilai gain tertinggi adalah mulai dari

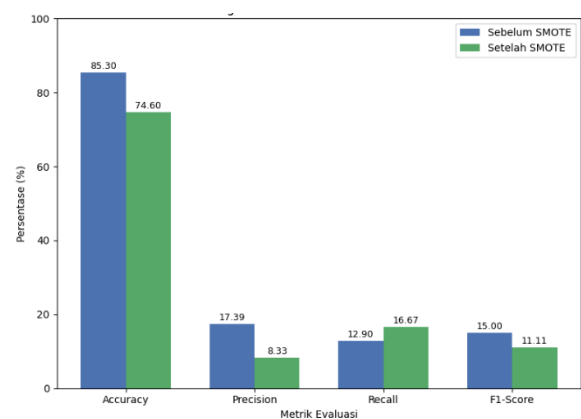
IP5, IP3, IP2, Jenis MBKM, Rata-rata nilai, Status pekerjaan, dan Usia mahasiswa.

Hasil percobaan naive bayes menggunakan rapidminer setelah dilakukan seleksi fitur dapat dilihat pada gambar 3.2. Hasil pengujian menunjukkan bahwa meskipun akurasi model Naïve Bayes mencapai 85,30%, performa terhadap kelas minoritas sangat rendah. Dari 32 mahasiswa yang tidak lulus tepat waktu, hanya 4 terprediksi benar, menghasilkan precision 17,39% dan recall 12,90%, sedangkan mayoritas (lulus tepat waktu) terklasifikasi hampir sempurna. Nilai AUC sebesar 0,500 mengindikasikan kemampuan diskriminasi model yang lemah akibat distribusi data yang tidak seimbang, sehingga akurasi tinggi tidak mencerminkan kinerja yang adil antar kelas.



Gambar 5 Akurasi Naive bayes

Untuk mengatasi masalah ini, dilakukan penerapan SMOTE (Synthetic Minority Oversampling Technique) pada data pelatihan. SMOTE bekerja dengan membuat sampel sintetis untuk kelas minoritas berdasarkan interpolasi antar titik data yang ada, sehingga distribusi kelas menjadi lebih seimbang tanpa mengurangi data mayoritas.



Gambar 6 Perbandingan Integrasi SMOTE

Setelah SMOTE, distribusi kelas menjadi seimbang pada data pelatihan, tetapi performa pada data uji menurun dibanding sebelumnya (akurasi turun dari 85.3% ke 74.6%). Recall meningkat (dari 12.9% ke 16.67%), artinya model sedikit lebih baik mengenali kelas minoritas. Precision tetap rendah,

menunjukkan prediksi kelas minoritas masih banyak salah. Penurunan akurasi wajar karena model kini mencoba mengenali kedua kelas, bukan hanya mayoritas.

Penelitian ini menghasilkan sebuah sistem prediksi kelulusan mahasiswa berbasis kecerdasan buatan yang diberi nama SMART GRADUATION (SmartGrad), dan telah diimplementasikan dalam bentuk aplikasi web interaktif yang dapat diakses melalui <https://smartgrad.streamlit.app>. Sistem ini merupakan penggabungan algoritma Decision Tree untuk seleksi fitur dan Naive Bayes untuk klasifikasi. Fitur-fitur penting seperti usia, IP semester, rata-rata nilai, jenis MBKM, dan status pekerjaan dipilih berdasarkan hasil seleksi Decision Tree. Model Naive Bayes kemudian digunakan untuk memprediksi apakah mahasiswa akan lulus tepat waktu atau tidak. Pendekatan learning analytics mencakup teknik seperti pemodelan prediktif, pemrosesan bahasa alami, dan analisis multimodal, yang semuanya memiliki relevansi tinggi untuk pengembangan SmartGrad[21].

Fitur utama SmartGrad: 1) Upload Data: Pengguna dapat mengunggah data mahasiswa baru. 2) Prediksi Kelulusan: Sistem menampilkan hasil prediksi apakah mahasiswa akan lulus tepat waktu atau tidak. 3) Visualisasi dan Interpretasi: Menampilkan probabilitas prediksi dan fitur yang paling berpengaruh, mendukung prinsip interpretable AI. Hasil pengujian menunjukkan bahwa sistem ini mampu memberikan prediksi yang akurat dan mudah dipahami. Dengan antarmuka yang sederhana, pengguna dapat memasukkan data mahasiswa dan memperoleh hasil prediksi secara instan. Sistem ini mendukung pendekatan interpretable AI, sehingga hasilnya dapat digunakan sebagai alat bantu pengambilan keputusan oleh dosen pembimbing maupun pengelola Kampus Merdeka. Penggunaan AI untuk mendukung kelulusan sangat penting diterapkan di perguruan tinggi[22]. Tampilan dashboard SmartGrad pada gambar 3.4.



Gambar 7 Dashboard SMARTGRAD

Pada halaman dashboard berisi fungsional yang dibutuhkan untuk melakukan prediksi kelulusan, dari

mulai input model yang diinginkan dari data set, kemudian dilakukan pelatihan dari data set tersebut, dan terdapat 7 fitur yang digunakan untuk melakukan prediksi dengan model naivebayes dari hasil seleksi fitur menggunakan decision tree.

Prediksi Kelulusan Tepat Waktu — SMART GRADUATION

Fitur dipakai: USIAMASUK, IP2, IP3, IP5, rata-rata nilai, mandiri/flagsip, BEKERJA/TIDAK Target (label): LULUS TEPAT/TIDAK

Data Pelatihan & Evaluasi Form Input (7 Fitur) Chatbot Tentang

2) Latih & Evaluasi (Fitur Terkunci)

Fitur dipakai: ["USIAMASUK", "IP2", "IP3", "IP5", "rata-rata nilai", "mandiri/flagsip", "BEKERJA/TIDAK"]

Nilai target positif

TEPAT

Latih Model Sekarang

Gambar 8 Fitur Model Pelatihan Data Training

Hasil akurasi prediksi yang dihasilkan oleh SMARTGRAD ditunjukkan pada gambar 3.6.

Nilai target positif
TEPAT

Latih Model Sekarang

Cross-Validation Aktif (10 folds)

Hasil Rata-rata Cross-Validation (10 folds)

	Metrik	Rata-rata	Standar Deviasi
0	Accuracy	0.888	± 0.022
1	Precision	0.905	± 0.016
2	Recall	0.979	± 0.024
3	F1-Score	0.940	± 0.012

Gambar 8 Tingkat akurasi yang dihasilkan

Hasil evaluasi aplikasi SmartGrad menunjukkan performa model yang sangat baik:

- Akurasi rata-rata mencapai 0,888 ($\pm 0,022$), menandakan bahwa sekitar 88% prediksi benar.
- Precision sebesar 0,905 ($\pm 0,016$), artinya sebagian besar prediksi “TEPAT” memang benar.
- Recall sangat tinggi, yaitu 0,979 ($\pm 0,024$), menunjukkan model hampir selalu berhasil mengenali mahasiswa yang lulus tepat waktu.
- F1-Score berada di 0,940 ($\pm 0,012$), mengindikasikan keseimbangan yang baik antara precision dan recall.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem prediksi kelulusan mahasiswa berbasis kecerdasan buatan yang diberi nama SmartGrad, dengan

mengintegrasikan algoritma Decision Tree untuk seleksi fitur dan Naïve Bayes untuk klasifikasi. Sistem ini mampu memberikan prediksi kelulusan tepat waktu secara akurat dan interpretatif, serta telah diimplementasikan dalam bentuk aplikasi web interaktif berbasis Streamlit, yang dapat diakses melalui <https://smartgrad.streamlit.app>. SmartGrad membuktikan bahwa pendekatan kombinasi algoritma dalam kecerdasan buatan dapat diterapkan secara praktis dalam dunia pendidikan tinggi, khususnya untuk mendukung pengambilan keputusan berbasis data dalam program Merdeka Belajar Kampus Merdeka. Hasil evaluasi menunjukkan performa model yang sangat baik dengan akurasi 88,8%, precision 90,5%, recall 97,9%, dan F1-score 94,0%, sehingga sistem ini layak digunakan sebagai alat bantu monitoring akademik.

Pengembangan selanjutnya dapat menguji algoritma lain seperti Random Forest, XGBoost, atau Deep Learning untuk meningkatkan akurasi. Menambahkan fitur baru, misalnya data kehadiran, aktivitas pembelajaran diluar kampus, dan indikator soft skills, agar sistem SmartGrad menjadi lebih komprehensif dalam mendukung keberhasilan studi mahasiswa.

5. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Ngudi Waluyo dan Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) UNW atas dukungan dalam pelaksanaan penelitian ini. Penelitian ini didanai oleh Direktorat Penelitian dan Pengabdian kepada Masyarakat (DPPM) melalui skema Penelitian Dosen Pemula (PDP) tahun 2025, yang telah memberikan dukungan sumber daya sehingga penelitian ini dapat diselesaikan dengan baik. Ucapan terima kasih juga disampaikan kepada koordinator akademik dan program Merdeka Belajar Kampus Merdeka yang telah memfasilitasi pengumpulan data serta semua pihak yang berkontribusi dalam pengembangan sistem SmartGrad.

DAFTAR PUSTAKA

- [1] A. Wibowo, D. Manongga, and H. D. Purnomo, "The utilization of naive bayes and C. 45 in predicting the timeliness of students' graduation," *Sci. J. Informatics*, vol. 7, no. 1, pp. 99–112, 2020.
- [2] D. Ifenthaler and D. Gibson, *Adoption of Data Analytics in Higher Education Learning and Teaching*. Springer, 2020.
- [3] Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, *Buku panduan Merdeka Belajar – Kampus Merdeka (Edisi II)*. 2024. [Online]. Available: <https://dikti.kemdikbud.go.id/wp-content/uploads/2024/06/Buku-Panduan-Merdeka-Belajar-Kampus-Merdeka-MBKM-2024.pdf>
- [4] TEM Journal, "Predicting Student Graduation Outcomes: An Evaluation of Project-Based Learning and Implementation of Naïve Bayes," *TEM J.*, vol. 14, no. 2, 2025.
- [5] International Journal of Computing and Digital Systems, "Comparative Analysis of Naïve Bayes and KNN Approaches to Predict Timely Graduation Using Academic History," *Int. J. Comput. Digit. Syst.*, vol. 16, no. 1, 2024.
- [6] AIP Conference Proceedings, "Predicting Graduate on Time Using Machine Learning," in *AIP Conference Proceedings*, 2024.
- [7] M. F. Faradillah and I. Saluza, "Perbandingan Model Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree dan Naive Bayes," *Decod. J. Pendidik. Teknol. Inf.*, 2025.
- [8] A. Wahyudi, Kusriani, and F. W. Wibowo, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Naïve Bayes," *J. Permata Indones.*, vol. 14, no. 2, pp. 132–138, 2023.
- [9] M. Julkarnain and Y. Yustiardin, "Penerapan Algoritma Naïve Bayes untuk Prediksi Kelulusan Mahasiswa," *J. Teknol. Inf. dan Komput.*, 2024.
- [10] I. Riadi, "Prediksi Kelulusan Tepat Waktu Berdasarkan Riwayat Akademik Menggunakan Metode K Nearest Neighbor," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 2, pp. 249–256, 2024.
- [11] M. Sivasakthi and K. R. A. Padmanabhan, "Prediction of Students Programming Performance Using Naïve Bayesian and Decision Tree," in *Soft Computing for Security Applications*, Springer, 2022, pp. 97–106.

- [12] MUTHIAH RIANI HARAHAP, “SISTEM PREDIKSI KELULUSAN MAHASISWA PADA PROGRAM MBKM MENGGUNAKAN METODE FUZZY MAMDANI,” 2024.
- [13] H. Ermamtita, “Prediksi Kelulusan Mahasiswa Menggunakan Metode K-Means dan Random Forest,” *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 1, pp. 209–214, 2025.
- [14] L. Setiyani, M. Wahidin, and D. Awaludin, “Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Systematic Review,” *Fakt. Exakta*, vol. 13, no. 1, pp. 35–43, 2020.
- [15] V. G. Costa and C. E. Pedreira, “Recent advances in decision trees: an updated survey,” *Artif. Intell. Rev.*, vol. 56, pp. 4765–4800, 2023, doi: 10.1007/s10462-022-10275-5.
- [16] P. Singh and B. Singh, “A Study of Data Mining Techniques for Student Performance Prediction,” Madurai, India: in 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 227–232.
- [17] M. A. A. M. M. N. Adnan, A. F. R. A. Rahman, N. A. H. A. Hamid, “Predicting student’s academic performance using data mining techniques: A systematic literature review,” *IEEE Access*, vol. 10, pp. 36723–36737, 2022, doi: 10.1109/ACCESS.2022.3163737.
- [18] and R. A. R. A. M. Salinitri, M. R. D. Girolamo, A. B. Petrova, “The impact of co-operative education on timely graduation: a competing risks analysis,” *High. Educ.*, vol. 83, no. 5, pp. 1033–1050, 2022, doi: 10.1007/s10734-021-00746-6.
- [19] and T. T. T. T. L. T. Vo, H. T. T. Nguyen, “The impact of scholarships on academic performance: a case study in Vietnam,” *Heliyon*, vol. 8, no. 9, p. e10577, 2022, doi: 10.1016/j.heliyon.2022.e10577.
- [20] D. Ifenthaler, D.-K. Mah, dan J. Yin-Kim Yau, *Utilizing Learning Analytics to Support Study Success*. Springer, 2019
- [21] C. Lang, G. Siemens, A. F. Wise, D. Gašević, dan A. Merceron (Eds.), *Handbook of Learning Analytics*, 2nd ed. SoLAR, 2022.