



## MENINGKATKAN AKURASI PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN METODE ALGORITMA GENETIKA

Sumarna<sup>1</sup>, Imam Nawawi<sup>2</sup>, Suhardjono<sup>3</sup>, Hari Sugiarto<sup>4</sup>, Dewi Yuliandari<sup>5</sup>

<sup>1</sup>Universitas Nusa Mandiri

<sup>2,3,4,5</sup>Universitas Bina Sarana Informatika

<sup>1</sup>Jl. Jatiwaringin Raya No. 02,

Makasar, Jakarta Timur 13620.

<sup>2,3,4,5</sup>Jl. Kramat Raya No. 98, Senen,

Jakarta Pusat 10450.

Email : [imamnawawi@bsi.ac.id](mailto:imamnawawi@bsi.ac.id)<sup>1</sup>, [sumarna.smn@nusamandiri.ac.id](mailto:sumarna.smn@nusamandiri.ac.id)<sup>2</sup>, [suhardjono@bsi.ac.id](mailto:suhardjono@bsi.ac.id)<sup>3</sup>, [hari.hrs@bsi.ac.id](mailto:hari.hrs@bsi.ac.id)<sup>4</sup>, [dewi.dwy@bsi.ac.id](mailto:dewi.dwy@bsi.ac.id)<sup>5</sup>

### ABSTRAK

Penelitian ini bertujuan untuk meningkatkan akurasi prediksi kelulusan mahasiswa melalui penerapan dan optimasi algoritma Naïve Bayes (NB) menggunakan Genetic Algorithm (GA). Teknik data mining digunakan untuk menganalisis atribut akademik dan demografis mahasiswa guna mengidentifikasi pola yang dapat memprediksi kemungkinan kelulusan mahasiswa. Penelitian ini menerapkan metode kuantitatif dengan memanfaatkan dataset berisi 796 catatan mahasiswa dari sebuah institusi pendidikan tinggi di Indonesia. Proses preprocessing data dilakukan untuk menangani missing value sebelum diterapkan ke dalam algoritma. Hasil algoritma NB menunjukan akurasi prediksi sebesar 83,30% dengan precision 57,91% dan recall 74,01%. Setelah dilakukan optimasi dengan GA, akurasi meningkat menjadi 85,43% dengan precision 67,10%, meskipun recall sedikit menurun menjadi 61,07%. Analisis perbandingan menunjukkan NB+GA lebih unggul dalam hal akurasi dan precision, namun algoritma NB tanpa optimasi memiliki keunggulan dalam recall yang penting untuk mendeteksi seluruh kasus positif. Ini menunjukkan bahwa integrasi GA dalam NB dapat meningkatkan beberapa aspek kinerja model, namun menimbulkan trade-off antara precision dan recall. Kombinasi NB dan GA tetap menawarkan pendekatan yang kompetitif dalam prediksi kelulusan mahasiswa. Hasil penelitian ini dapat menjadi dasar untuk pengembangan sistem pendukung keputusan di lembaga pendidikan tinggi.

**Kata kunci :** Genetic Algorithm, Meningkatkan Akurasi, Naïve Bayes, Prediksi Kelulusan Mahasiswa.

### ABSTRACT

This research aims to improve the accuracy of student graduation predictions through the implementation and optimization of the Naïve Bayes (NB) algorithm using Genetic Algorithm (GA). Data mining techniques are used to analyze academic and demographic attributes of students to identify patterns that can predict the likelihood of graduation. This study applies a quantitative method utilizing a dataset of 796 student records from a higher education institution in Indonesia. Data preprocessing is carried out to handle missing values before applying the algorithm. The results of the NB algorithm show a prediction accuracy of 83.30% with a precision of 57.91% and a recall of 74.01%. After optimization with GA, the accuracy increased to 85.43%, with a precision of 67.10%, although the recall slightly decreased to 61.07%. Comparative analysis shows that NB+GA is superior in terms of accuracy and precision, while the unoptimized NB algorithm has an advantage in recall, which is important for detecting all positive cases. This indicates that the integration of GA into NB can improve some aspects of model performance but creates a trade-off between precision and recall. The combination of NB and GA still offers a competitive approach in predicting student graduation. The results of this research can serve as a basis for the development of decision support systems in higher education institutions.

**Kata kunci :** Genetic Algorithm, Improving Accuracy, Naïve Bayes, Student Graduation Prediction.



## 1. PENDAHULUAN

Perguruan tinggi merupakan sebagai lembaga institusi pendidikan tinggi yang memberikan pendidikan akademik dan profesional kepada mahasiswa. Mahasiswa sebagai individu yang sedang menjalani studi akademik di perguruan tinggi dari masuk sampai lulus memiliki data pribadi masing-masing yang tersimpan dalam sebuah database yang dikelola oleh pihak kampus tersebut. Data mengenai mahasiswa meliputi informasi pribadi, riwayat akademik, partisipasi dalam kegiatan ekstrakurikuler, dan informasi lain yang relevan (Laurentinus et al., 2022). Sebagai lembaga pendidikan perguruan tinggi mengumpulkan jumlah data yang besar dan menyimpan data mahasiswa termasuk data akademik, sosial, administratif, kehadiran, dan faktor lainnya (Salam, Zeniarja, & Anthareza, 2022). Data yang besar sangat sulit dicari informasi dari mahasiswa tersebut. Maka dari itu diperlukan teknik datamining. Penggunaan teknik datamining, seperti klasifikasi dan regresi semakin menjadi fokus untuk menganalisis data mahasiswa guna mengidentifikasi pola yang dapat memprediksi dengan lebih tepat kemungkinan kelulusan mahasiswa (Bokgoshi, Sixhaxa, Jadhav, Nyamane, & Ajoodha, 2023). Data Mining dengan cara kerja memproses ekstraksi pola yang tersembunyi dan memberikan pengetahuan yang bermanfaat yang diperoleh dari data yang besar dan kompleks (Gustiana & Sari, 2021). Teknik ini mencakup berbagai metode seperti prediksi, *clustering*, klasifikasi, regresi, dan asosiasi untuk menganalisis data dan mengidentifikasi pola yang bermanfaat. Pada penelitian ini digunakan dengan teknik prediksi terhadap kelulusan mahasiswa (Lin, 2023). Kelulusan mahasiswa merupakan pencapaian akademik yang penting bagi perguruan tinggi dan mahasiswa itu sendiri (Liao, 2023). Prediksi kelulusan mahasiswa dapat membantu dalam menyediakan bimbingan akademik yang tepat, mengoptimalkan proses akademik, serta mengurangi tingkat *drop-out* atau keterlambatan dalam menyelesaikan studi (Kurniawati & Maulidevi, 2022). Prediksi kelulusan mahasiswa berdasarkan data historis mereka selama masa studi menjadi krusial karena dapat mempengaruhi strategi manajemen akademik dan pengambilan keputusan di perguruan tinggi, serta membantu dalam meningkatkan retensi mahasiswa dan efisiensi program pendidikan (Ong & Pedrasa, 2021). Pada banyak lembaga pendidikan tinggi, prediksi kelulusan mahasiswa merupakan hal yang penting untuk membantu dalam perencanaan dan pengambilan keputusan (Karmagatri, Kurnianingrum, Suciana, & Utami, 2023). Dengan

memiliki prediksi yang tepat, universitas dapat mengalokasikan sumber daya secara lebih efisien, memberikan bimbingan yang lebih baik kepada mahasiswa, serta meningkatkan tingkat retensi dan kelulusan (Rawatlal, Chetty, & Naicker, 2022).

Tujuan dari penelitian ini adalah meningkatkan akurasi prediksi kelulusan mahasiswa dengan menggunakan metode Algoritma Genetika. Dengan meningkatkan akurasi prediksi, diharapkan universitas dapat lebih efisien dalam mengelola sumber daya dan memberikan dukungan yang lebih baik kepada mahasiswa.

Beberapa penelitian sebelumnya telah menerapkan teknik data mining untuk melakukan prediksi kelulusan mahasiswa dengan data yang sama untuk digunakan dalam penelitian DT, SVM, dan NN (Riyanto, Hamid, & Ridwansyah, 2019), DT dan PSO (Hendra, Azis, & Suhardjono, 2020), SVM dan PSO (Suhardjono, Wijaya, & Hamid, 2019), SVM dan GA (Ridwansyah, Wijaya, & Purnama, 2020), NN dan PSO (Nurdin, Sartini, Sumarna, Maulana, & Riyanto, 2023). Pada tahun 2024 telah dilaksanakan studi tentang deteksi kelulusan pada mahasiswa dengan menggunakan Algoritma Genetika (GA) dengan hasil evaluasi kinerja algoritma genetika menggunakan nilai akurasi sebesar 87.33% (Pangesti, Ariyati, Priyono, Sugiono, & Suryadithia, 2024) dan 87.68% (Rukiastindari, Rohimah, Aprillia, & Mutia, 2024). Penelitian lain dengan menggunakan data yang sama dan tahun yang sama menggunakan metode Naïve Bayes (NB) untuk prediksi kelulusan mahasiswa memperoleh hasil akurasi sebesar 83.30% (Kahfi, Prihatin, Yudhistira, Sudradjat, & Wijaya, 2024).

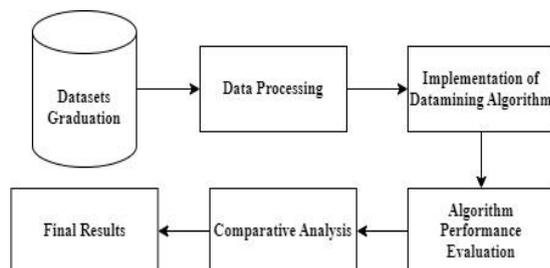
Kontribusi penelitian ini yaitu melakukan studi perbandingan menggunakan algoritma data mining yang sebelumnya telah diterapkan yaitu *Naïve Bayes* dan *Genetic Algorithm*, dengan menguji masing-masing algoritma tersebut secara terpisah untuk memprediksi kelulusan mahasiswa. Sebagai pembeda dari studi perbandingan sebelumnya, penelitian ini menggabungkan kedua metode tersebut untuk pengujian data kelulusan mahasiswa. Tujuannya adalah meningkatkan akurasi metode *Naïve Bayes* melalui penggunaan *Genetic Algorithm* sebagai metode pengoptimalan. Sehingga penelitian ini bertujuan meningkatkan kinerja akurasi algoritma *Naïve Bayes*, yang belum dilakukan pada penelitian sebelumnya yang membandingkan kedua algoritma tersebut. Dengan demikian, penelitian ini diharapkan dapat berkontribusi dalam meningkatkan efektivitas prediksi kelulusan mahasiswa menggunakan pendekatan *Genetic Algorithm*, yang dapat menjadi dasar untuk pengembangan sistem pendukung



keputusan di lembaga pendidikan tinggi.

## 2. METODE PENELITIAN

Dalam melakukan penelitian untuk memprediksi kelulusan mahasiswa dilakukan jenis penelitian kuantitatif yang digunakan, yaitu jenis penelitian ilmiah yang dilakukan secara sistematis untuk mempelajari bagian-bagian dari suatu fenomena serta kualitas hubungan antar bagian tersebut. Tujuan penelitian kuantitatif untuk mengukur dan menganalisis variabel-variabel yang terlibat dalam fenomena tertentu secara sistematis dan objektif, guna menghasilkan data statistik yang dapat digunakan untuk membuat prediksi kelulusan mahasiswa secara tepat waktu. Penelitian ini berfokus pada peningkatan metode *Naive Bayes* dengan algoritma optimasi berbasis *Genetic Algorithm* untuk memprediksi apakah mahasiswa lulus tepat waktu atau tidak. Dalam melakukan prediksi kelulusan, maka dilakukan ilustrasi visual dari tahapan dan langkah-langkah yang akan dikerjakan dengan diagram alir penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Flow Diagram

### 2.1 Dataset

Dataset yang dilakukan untuk memprediksi kelulusan mahasiswa merupakan dataset primer yang diambil dari salah satu institusi perguruan tinggi swasta yang ada di Indonesia yaitu dataset kelulusan dengan total data 796 *record*. Atribut yang terdapat pada dataset tersebut terdiri dari jenis kelamin, jurusan SLTA, sekolah asal SLTA, dan tepat waktu. Atribut jenis kelamin dengan nilai *record*-nya adalah 1 dan 2. Atribut jurusan SLTA dengan nilai *record*-nya 1 sampai 10 dimana ada *record* jurusan IPA, IPS, Teknik Komputer Jaringan, Multimedia, Informatika dan sebagainya. Atribut sekolah SLTA dengan nilai *record* nya 1 sampai 5 dimana ada *record* asal SMA Negeri, SMA Swasta, SMK Negeri, SMK Swasta, Madrasah Aliyah Negeri dan Madrasah Aliyah Swasta. Atribut indeks prestasi kumulatif dengan nilai *record* 0,00 sampai 4,00. Dan terakhir hasil dari atribut tersebut

yaitu tepat waktu dengan nilai *record* ya atau tidak.

### 2.2 Data Processing

Tahap pengelolaan data yang dilakukan penelitian tentang prediksi kelulusan mahasiswa dilakukan dengan menggunakan proses imputasi rata-rata, yaitu suatu metode untuk mengatasi data yang tidak tersedia pada dataset kelulusan mahasiswa tersebut. Proses pengerjaan imputasi tersebut dilakukan dengan nilai yang hilang pada suatu atribut diganti dengan nilai rata-rata atribut tersebut (Yennimar, Leonardi, Weide, Cantona, & Gani Mores Hutagalung, 2024). Metode proses Imputasi tersebut berperan dalam mempertahankan konsistensi, keakuratan dan keandalan terhadap dataset kelulusan mahasiswa tersebut yang dilakukan secara langsung dengan menggunakan program *Rapid Miner 5.0*.

### 2.3 Implementation of Datamining Algorithm

Pada tahap implementasi dengan algoritma datamining, dilakukan implementasi menggunakan dua algoritma data mining yang akan dianalisis dalam penelitian ini yaitu algoritma *Naive Bayes*, dan *Naive Bayes* berbasis *Genetic Algorithm* untuk memprediksi kelulusan tepat waktu atau tidak pada mahasiswa yang ada di institusi tersebut. Algoritma tersebut diimplementasikan menggunakan aplikasi *rapid miner 5.0*.

### 2.4 Algorithm Performance Evaluation

Kinerja algoritma dalam penelitian prediksi kelulusan mahasiswa dievaluasi menggunakan *Confusion Matrix*, yaitu sebuah tabel silang yang mengelompokkan data kelas positif dan data kelas negatif ke dalam kategori. Kategori tersebut dibagi menjadi dua kategori, dimana terdapat kategori prediksi dan kategori aktual. Kategori prediksi memiliki hasil berdasarkan data yang dimasukkan kedalam dataset, Kategori aktual dengan hasil yang sebenarnya atau yang benar berdasarkan data yang telah diverifikasi atau diketahui.

### 2.5 Comparative Analysis

Analisis komparatif dilakukan untuk membandingkan hasil dari prediksi yang dihasilkan oleh kedua algoritma tersebut. *Naive Bayes* digunakan sebagai algoritma dasar, dan hasilnya dibandingkan dengan *Naive Bayes* yang telah dioptimalkan dengan *Genetic Algorithm*. Tujuan dari analisis ini adalah untuk menentukan sejauh mana *Genetic Algorithm* dapat meningkatkan akurasi prediksi *Naive Bayes*. Hasil analisis diharapkan metode *Naive Bayes* yang dioptimalkan dengan *Genetic Algorithm* memiliki performa yang lebih baik dalam memprediksi



kelulusan mahasiswa dibandingkan dengan Naïve Bayes tanpa optimasi. Hal ini yang nantinya akan ditunjukkan oleh peningkatan nilai metrik evaluasi seperti akurasi dan kurva ROC. Kurva ROC menggambarkan trade-off antara True Positive Rate (Sensitivitas) dan False Positive Rate (1 - Spesifisitas) ketika ambang keputusan (threshold) bervariasi (Islam, Banik, Rahman, & Rahman, 2023).

**2.6 Final Results**

Hasil penelitian ini disajikan dalam bentuk pembahasan yang menyeluruh dan terperinci, yang menguraikan seluruh analisis yang telah dilakukan dengan sangat mendetail. Setiap langkah dalam analisis ini dijelaskan secara rinci untuk memberikan pemahaman yang mendalam tentang proses dan hasil yang diperoleh. Pembahasan ini tidak hanya berfokus pada interpretasi data dan hasil akhir, tetapi juga memperhatikan bagaimana analisis ini berkontribusi pada pengetahuan yang ada dengan menghubungkannya secara langsung dengan penelitian-penelitian sebelumnya yang relevan.

Dalam pembahasan ini, hasil penelitian dibandingkan dan dikontraskan dengan temuan dari studi-studi terdahulu untuk menilai kesesuaian dan inovasi yang ditawarkan oleh penelitian ini. Dengan mengaitkan hasil penelitian ini dengan literatur yang ada, peneliti berupaya menunjukkan konsistensi atau perbedaan yang signifikan dengan studi-studi sebelumnya, serta mengidentifikasi kontribusi unik yang diberikan oleh penelitian ini terhadap perkembangan pengetahuan di bidang yang diteliti. Pembahasan ini bertujuan untuk memberikan gambaran yang komprehensif tentang bagaimana penelitian ini menambah nilai dan memperkaya pemahaman tentang topik yang sedang diteliti, serta memberikan wawasan baru yang dapat menjadi dasar bagi penelitian di masa mendatang.

**3. HASIL DAN PEMBAHASAN**

Hasil pengujian dalam penelitian ini dianalisis menggunakan Rapid Miner 5.0, dengan tujuan untuk mengevaluasi kinerja algoritma *Naïve Bayes*, dan *Naïve Bayes* berbasis *Genethic Algorithm* dalam memprediksi kelulusan mahasiswa secara tepat waktu.

**3.1 Dataset**

Gambar 2 merupakan kumpulan dataset yang akan dilakukan pengujian dalam penelitian prediksi kelulusan mahasiswa ini

JK	JRS	Asal	IPK1	IPK2	IPK 3	IPK4	IPK5	IPK6	Tepat
1	9	5	3,09	2,88	2,89	3,02	3,1	3,15	Ya
1	7	3	2,86	2,88	3,1	3,11	3,14	3,16	Ya
	7	5	2,55	2,56	2,66	2,63	2,73	2,96	Tidak
1		3	3,41	3,05	3,07	2,79	2,96	3,09	Tidak
	7	3	3,09	3,07	3,02	3,07	3,14	3,28	Ya
1	5	3	2,32	2,41	2,49	2,52	2,91	2,99	Ya
1	5	2	1,41	1,46	1,74	1,6	1,97	2,61	Tidak
	5	4	2,45	2,63	2,41	2,58	2,95	2,95	Tidak
...	...	...	...	...	...	...	...	...	...
2	4	3	2,18	2,21	2,37	2,64	2,91	3,01	Ya
2	1		3,41	3,4	3,4	3,48	3,5	3,52	Ya
2	7	3	1,95	2,55	2,58	2,53	2,43	2,6	Tidak

Gambar 2. Dataset

Pada Gambar 2 untuk keterangan JK menerangkan jenis kelamin dimana 1 merupakan laki-laki 2 merupakan perempuan. JRS menerangkan jurusan asal slta dimana menerangkan jurusan SLTA dari calon mahasiswa. asal menerangkan asal SLTA dimana asal SLTA dari calon mahasiswa. IPK menerangkan indeks prestasi kumulatif dari semester 1 sampai semester 6. tepat menerangkan tepat waktu dimana mahasiswa mengambil studi dengan kelulusan tepat waktu atau tidak.

**3.2 Data Processing**

Dalam dataset yang digunakan dalam penelitian ini, masih terdapat beberapa data yang tidak memiliki nilai atau yang dikenal sebagai missing value, sehingga diperlukan tahap *preprocessing* pada dataset tersebut. Data setelah tahap *preprocessing* dapat dilihat pada Gambar 3.

JK	JRS	Asal	IPK1	IPK2	IPK 3	IPK4	IPK5	IPK6	Tepat
1	9	5	3,09	2,88	2,89	3,02	3,1	3,15	Ya
1	7	3	2,86	2,88	3,1	3,11	3,14	3,16	Ya
1	7	5	2,55	2,56	2,66	2,63	2,73	2,96	Tidak
1	7	3	3,41	3,05	3,07	2,79	2,96	3,09	Tidak
1	7	3	3,09	3,07	3,02	3,07	3,14	3,28	Ya
1	5	3	2,32	2,41	2,49	2,52	2,91	2,99	Ya
1	5	2	1,41	1,46	1,74	1,6	1,97	2,61	Tidak
1	5	4	2,45	2,63	2,41	2,58	2,95	2,95	Tidak
...	...	...	...	...	...	...	...	...	...
2	4	3	2,18	2,21	2,37	2,64	2,91	3,01	Ya
2	1	3	3,41	3,4	3,4	3,48	3,5	3,52	Ya
1	9	5	3,09	2,88	2,89	3,02	3,1	3,15	Tidak

Gambar 3. Dataset Setelah *Processing*



Pada Gambar 3 yang merupakan hasil dari data setelah di proses untuk missing value dengan menggunakan aplikasi rapid miner. Proses preprocessing ini menghilangkan atau mengisi nilai yang hilang dengan menggunakan metode yang sesuai untuk memastikan bahwa dataset siap untuk analisis lebih lanjut dengan menggunakan metode naïve bayes. Dengan data yang telah diproses dengan baik, analisis yang dilakukan selanjutnya akan lebih akurat dan dapat memberikan wawasan yang lebih berguna mengenai prediksi kelulusan mahasiswa.

**3.3 Algorithm Performance Evaluation**

Pada tahap ini, algoritma datamining diterapkan untuk memprediksi kelulusan mahasiswa menggunakan dua pendekatan utama, yaitu NB dan NB yang dioptimalkan dengan GA. Proses implementasi data yang sudah dilakukan pada tahap *preprocessing* dilakukan dengan menggunakan aplikasi *Rapid Miner 5.0*, yang menyediakan berbagai alat dan teknik untuk analisis data yang komprehensif. Hasil dari data olahan tersebut dengan *naïve bayes* dengan dapat dilihat pada tabel 1.

Tabel 1. Distribusi Tabel

Attribute	Parameter	YA	TIDAK
JK	mean	1.421	1.398
JK	standard dev	0.494	0.449
JRS	mean	5.689	5.804
JRS	standard dev	2.517	2.633
Asal	mean	2.899	3.111
Asal	standard dev	1.015	0.965
IPK1	mean	2.869	2.390
IPK1	standard dev	0.442	0.539
IPK2	mean	2.900	2.342
IPK2	standard dev	0.440	0.526
IPK3	mean	2.945	2.339
IPK3	standard dev	0.414	0.497
IPK4	mean	3.051	2.405
IPK4	standard dev	0.370	0.478
IPK5	mean	3.166	2.603
IPK5	standard dev	0.296	0.398
IPK6	mean	3.204	2.801
IPK6	standard dev	0.271	0.280

Tabel 1 menggambarkan distribusi statistik dari beberapa atribut yang digunakan dalam model NB untuk memprediksi kelulusan mahasiswa (YA dan TIDAK). Atribut-atribut ini mencakup variabel demografis dan akademik, seperti jenis kelamin (JK), jurusan SLTA (JRS), asal SLTA (Asal), dan indeks prestasi kumulatif dari semester 1 hingga semester 6 (IPK1 hingga IPK6). Berikut keterangan dari tabel 3.

a. Atribut JK (Jenis Kelamin):

Mean: Rata-rata nilai untuk jenis kelamin menunjukkan bahwa kelompok YA (mahasiswa yang lulus tepat waktu) memiliki rata-rata 1.421, sementara kelompok TIDAK (mahasiswa yang tidak lulus tepat waktu) memiliki rata-rata 1.398. Nilai ini mungkin dikodekan sebagai 1 untuk laki-laki dan 2 untuk perempuan.

Standard Dev: Deviasi standar menunjukkan seberapa tersebar data dalam kelompok tersebut. Dalam hal ini, kedua kelompok memiliki deviasi standar yang hampir sama, menunjukkan bahwa variasi dalam atribut jenis kelamin cukup seragam di antara kedua kelompok.

b. Atribut JRS (Jurusan SLTA):

Mean: Rata-rata untuk jurusan SLTA sedikit berbeda antara kelompok YA (5.689) dan TIDAK (5.804). Ini menunjukkan perbedaan kecil dalam distribusi jurusan SLTA antara kedua kelompok.

Standard Dev: Deviasi standar untuk jurusan SLTA juga sedikit lebih tinggi pada kelompok TIDAK, menunjukkan variasi yang sedikit lebih besar dalam latar belakang akademik SLTA dari mahasiswa yang tidak lulus tepat waktu.

c. Atribut Asal (Asal SLTA):

Mean: Rata-rata untuk asal SLTA menunjukkan bahwa mahasiswa yang lulus tepat waktu cenderung memiliki nilai rata-rata 2.899, sementara yang tidak lulus tepat waktu memiliki rata-rata 3.111.

Standard Dev: Deviasi standar untuk asal SLTA hampir sama antara kedua kelompok, yang menunjukkan bahwa sebaran asal SLTA mahasiswa relatif seragam.

d. Atribut IPK (Indeks Prestasi Kumulatif):

Mean: IPK rata-rata pada kelompok YA secara konsisten lebih tinggi, misalnya pada IPK1 dengan rata-rata 2.869 untuk YA dan 2.390 untuk TIDAK. Pola ini terus berlanjut hingga semester 6 (IPK6), menunjukkan bahwa kinerja akademik mahasiswa secara keseluruhan lebih baik pada kelompok yang lulus tepat waktu.

Standard Dev: Deviasi standar untuk IPK juga lebih rendah pada kelompok YA dibandingkan dengan TIDAK, yang menunjukkan bahwa nilai IPK di antara mahasiswa yang lulus tepat waktu lebih konsisten dan kurang tersebar.

Tabel 1 menunjukkan rata-rata IPK dari semester 1 (IPK1) hingga semester 6 (IPK6). Terlihat bahwa mahasiswa yang lulus tepat waktu (YA) cenderung memiliki IPK yang lebih tinggi di semua semester dibandingkan dengan mahasiswa yang tidak lulus tepat waktu (TIDAK).

Dari tabel ini, dapat disimpulkan bahwa ada perbedaan yang jelas dalam karakteristik akademik antara mahasiswa yang lulus tepat waktu dan yang



tidak. Mahasiswa yang lulus tepat waktu cenderung memiliki nilai IPK yang lebih tinggi dan lebih konsisten sepanjang masa studi mereka. Faktor-faktor seperti jenis kelamin, jurusan SLTA, dan asal SLTA juga menunjukkan perbedaan kecil antara kedua kelompok, namun pengaruh terbesar tampaknya berasal dari kinerja akademik (IPK).

Setelah itu dilakukan implementasi metode NB berbasis GA dengan menunjukkan bobot dari masing-masing atribut yang digunakan dalam model prediksi kelulusan mahasiswa. Bobot ini mencerminkan pentingnya setiap atribut dalam kontribusinya terhadap prediksi model kelulusan mahasiswa.

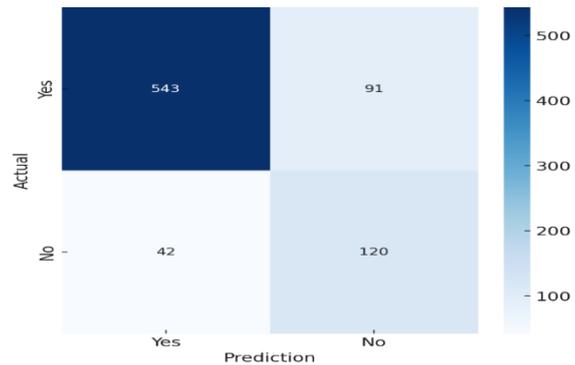
Tabel 2. Bobot Atribut

Attribute	Weight
JK	0
JRS	0
ASAL	0
IPK1	0
IPK2	0
IPK3	0
IPK4	0
IPK5	0
IPK6	1

Keterangan dari tabel 2 diketahui bahwa Atribut IPK6 memiliki bobot. ini berarti atribut IPK 3 dan IPK6 (Indeks Prestasi Kumulatif pada semester 1 dan 6) adalah satu-satunya atribut yang memiliki bobot yang tidak nol, menunjukkan bahwa atribut ini sangat penting dalam model prediksi. Bobot 1 menunjukkan bahwa atribut ini memiliki kontribusi penuh dalam memprediksi kelulusan mahasiswa tepat waktu. Sedangkan atribut yang lain memiliki bobot 0 hal ini menunjukkan bahwa model prediksi tidak menganggap atribut-atribut ini sebagai faktor yang signifikan.

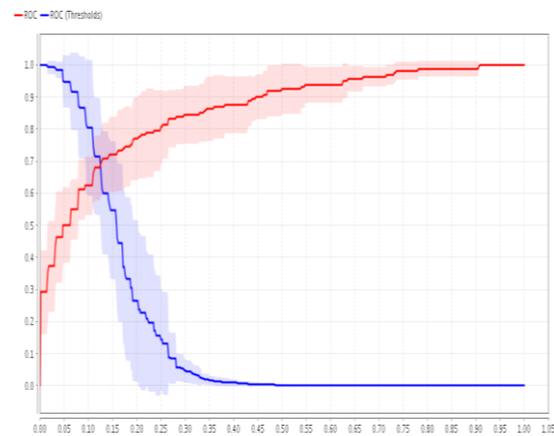
**3.4 Algorithm Performance Evaluation**

Pada tahap ini, kinerja algoritma dalam memprediksi kelulusan mahasiswa dievaluasi menggunakan *Confusion Matrix*. *Confusion Matrix* adalah alat evaluasi yang digunakan untuk menilai performa model klasifikasi dengan cara membandingkan prediksi yang dibuat oleh model dengan hasil aktual. Hasil *Confusion Matrix* yang didapat dengan menggunakan model NB pada Gambar 4.



Gambar 4. Confusion Matrix untuk NB

Pada Gambar 4 hasil evaluasi kinerja dengan NB *True Positive* (TP): 543 mahasiswa yang diprediksi akan lulus tepat waktu (YA) oleh model, dan benar-benar lulus tepat waktu. *True Negative* (TN): 120 mahasiswa yang diprediksi tidak akan lulus tepat waktu (TIDAK) oleh model, dan benar-benar tidak lulus tepat waktu. *False Positive* (FP): 91 mahasiswa yang diprediksi akan lulus tepat waktu (YA) oleh model, tetapi ternyata tidak lulus tepat waktu. *False Negative* (FN): 42 mahasiswa yang diprediksi tidak akan lulus tepat waktu (TIDAK) oleh model, tetapi ternyata lulus tepat waktu. Dari hasil tersebut dapat terbentuk *Area Under Curve* (AUC) dengan *Receiver Operating Characteristic* (ROC) yang terlihat Pada gambar 5.

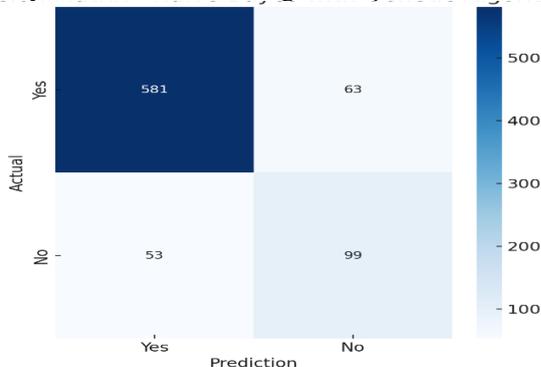


Gambar 5. AUC Model NB

Kurva ROC sering dievaluasi berdasarkan *Area Under the Curve* (AUC). Semakin besar nilai AUC, semakin baik kinerja model secara keseluruhan. Pada Gambar 3 AUC = 0,860 menunjukkan bahwa model memiliki kinerja yang sangat baik dalam membedakan antara kelas positif dan negatif. Nilai AUC mendekati 1 menunjukkan bahwa model ini hampir sempurna dalam prediksinya.

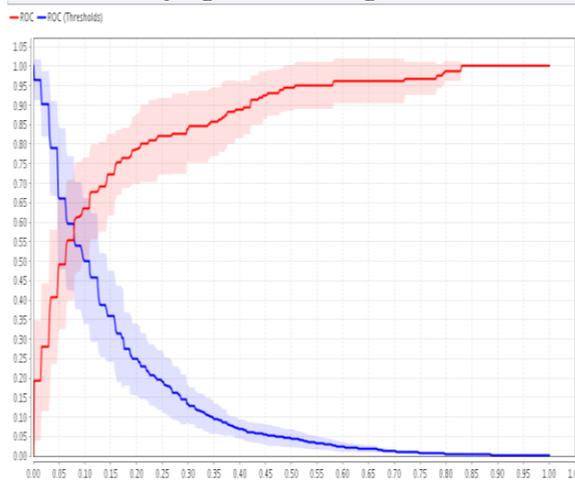


Setelah dilakukan dengan metode naïve bayes maka akan dilakukan peningkatan dengan metode genetic algorithm. Dari hasil didapat confusion matrix setelah dilakukan dengan metode Naive Bayes berbasis Genetic Algorithm pada Gambar 6.



Gambar 6. Confusion Matrix untuk NB dengan GA

Pada Gambar 6 hasil evaluasi kinerja dengan Naïve Bayes True Positive (TP): 581 mahasiswa yang diprediksi akan lulus tepat waktu (YA) oleh model, dan benar-benar lulus tepat waktu. True Negative (TN): 99 mahasiswa yang diprediksi tidak akan lulus tepat waktu (TIDAK) oleh model, dan benar-benar tidak lulus tepat waktu. False Positive (FP): 63 mahasiswa yang diprediksi akan lulus tepat waktu (YA) oleh model, tetapi ternyata tidak lulus tepat waktu. False Negative (FN): 53 mahasiswa yang diprediksi tidak akan lulus tepat waktu (TIDAK) oleh model, tetapi ternyata lulus tepat waktu. Dari hasil tersebut dapat terbentuk Area Under Curve (AUC) dengan Receiver Operating Characteristic yang terlihat Pada gambar 7.



Gambar 7. AUC Model NB with GA

Pada Gambar 5 AUC = 0,863 menunjukkan bahwa model memiliki kinerja yang sangat baik dalam membedakan antara kelas positif dan negatif. Nilai AUC mendekati 1 menunjukkan bahwa model ini

hampir sempurna dalam prediksinya.

### 3.5 Comparative Analysis

Setelah proses evaluasi kinerja algoritma dilakukan, langkah selanjutnya adalah membandingkan kedua algoritma yang diuji dalam penelitian ini untuk menentukan algoritma mana yang lebih efektif dalam memprediksi kelulusan mahasiswa. Hasil analisis perbandingan disajikan pada Tabel 3 di bawah ini.

Tabel 3. Analisis Perbandingan Algoritma

Algoritma	NB	NB+GA
<i>Accuracy</i>	83.30%	85.43%
<i>Precision</i>	57.91%	67.10%
<i>Recall</i>	74.01%	61.07%
<i>AUC</i>	0.860	0.863%

Algoritma NB memiliki akurasi sebesar 83.30%, sementara algoritma yang dioptimalkan dengan GA (NB+GA) menunjukkan peningkatan akurasi menjadi 85.43%. Hal ini menunjukkan bahwa NB+GA lebih baik dalam memprediksi kelulusan mahasiswa secara keseluruhan. Precision pada algoritma NB+GA adalah 67.10%, lebih tinggi dibandingkan dengan NB yang memiliki precision 57.91%. Ini menunjukkan bahwa NB+GA lebih andal dalam memberikan prediksi positif yang benar. Recall pada NB adalah 74.01%, lebih tinggi dibandingkan dengan NB+GA yang memiliki recall 61.07%. Hal ini menunjukkan bahwa algoritma NB lebih baik dalam mendeteksi semua kasus positif. Nilai AUC untuk NB adalah 0.860, sedangkan NB+GA menunjukkan sedikit peningkatan dengan AUC sebesar 0.863%. AUC yang lebih tinggi pada NB+GA menunjukkan kemampuan yang sedikit lebih baik dalam membedakan antara kelas positif dan negatif.

### 3.6 Final Results

Dari hasil perbandingan ini, dapat disimpulkan bahwa metode NB yang dioptimalkan dengan GA (NB+GA) memiliki kinerja yang lebih baik secara keseluruhan dalam hal akurasi dan precision. Namun dalam hal recall, algoritma NB tanpa optimasi memiliki keunggulan yang lebih baik dalam mendeteksi kasus-kasus positif.

Pengetahuan ini dapat digunakan untuk mengembangkan model prediksi yang lebih akurat dalam mengidentifikasi mahasiswa yang berpotensi tidak lulus tepat waktu, sehingga lembaga pendidikan dapat memberikan intervensi yang lebih dini dan tepat sasaran.



#### 4. KESIMPULAN

Penelitian ini berhasil melakukan evaluasi komprehensif terhadap kinerja dua algoritma, NB dan NB yang dioptimalkan dengan GA (NB+GA) dalam konteks prediksi kelulusan mahasiswa. Dari proses evaluasi tersebut, hasil menunjukkan bahwa kedua algoritma memiliki kapabilitas yang signifikan dalam memprediksi kelulusan mahasiswa, tetapi dengan perbedaan yang menonjol dalam beberapa aspek kinerja. Dengan hasil berdasarkan peningkatan nilai akurasi dari 83,30% menjadi 85,43% perbaikan nilai *precision* dari 57,91% menjadi 67,10%, stabilitas model kurva. Akan tetapi mendapatkan penurunan dari nilai recall.

Penelitian ini menyediakan bukti bahwa integrasi teknik GA dalam NB tidak hanya meningkatkan beberapa aspek kinerja tetapi juga menonjolkan trade-off antara *precision* dan *recall*, yang harus dipertimbangkan dalam penggunaan praktis. Ini relevan terutama dalam konteks pendidikan di mana penting untuk menyeimbangkan antara tidak melewatkan mahasiswa yang berpotensi lulus dengan mengidentifikasi secara akurat kelulusan mereka yang mungkin memerlukan intervensi lebih lanjut.

#### 5. REFERENSI

- Bokgoshi, L., Sixhaxa, K., Jadhav, A., Nyamane, S., & Ajoodha, R. (2023). Enhancing Timely Graduations: An Explainable AI Approach to Predict Academic Risks in South African Students. *International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE. <https://doi.org/10.1109/ICECET58911.2023.10389444>
- Gustiana, Z., & Sari, A. N. (2021). Combination of C 4.5 Algorithm and Profile Matching for Determining University Students Graduation. *International Conference on Computer Science and Engineering (IC2SE)*. IEEE. <https://doi.org/10.1109/IC2SE52832.2021.9791954>
- Hendra, Azis, M. A., & Suhardjono. (2020). ANALISIS PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN DECISION TREE BERBASIS PARTICLE SWARM OPTIMIZATION. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(1), 102–107. <https://doi.org/https://doi.org/10.32736/sisfokom.v9i1.756>
- Islam, M. R., Banik, S., Rahman, K. N., & Rahman, M. M. (2023). A comparative approach to alleviating the prevalence of diabetes mellitus using machine learning. *Computer Methods and Programs in Biomedicine*, 4. <https://doi.org/https://doi.org/10.1016/j.cmpbup.2023.100113>
- Kahfi, A. H., Prihatin, T., Yudhistira, Sudradjat, A., & Wijaya, G. (2024). THE RIGHT STEPS TOWARDS GRADUATION: NB-PSO SMART COMBINATION FOR STUDENT GRADUATION PREDICTION. *Jurnal Teknik Informatika (Jutif)*, 5(2), 607–614. <https://doi.org/https://doi.org/10.52436/1.jutif.2024.5.2.1889>
- Karmagatri, M., Kurnianingrum, D., Suciana, M. R., & Utami, S. A. (2023). Predicting Factors Related to Student Performance Using Decision Tree Algorithm. *International Conference on Cybernetics and Intelligent System (ICORIS)*. IEEE. <https://doi.org/10.1109/ICORIS60118.2023.10352269>
- Kurniawati, G., & Maulidevi, N. U. (2022). Multivariate Sequential Modelling for Student Performance and Graduation Prediction. *International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. IEEE. <https://doi.org/10.1109/ICITACEE55701.2022.9923971>
- Laurentinus, Rizan, O., Sarwindah, Hamidah, Sulaiman, R., & Fuston, P. (2022). Data Mining Using C4.5 Algorithm in Predicting Student Graduation. *International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. IEEE. <https://doi.org/10.1109/ISRITI56927.2022.10052793>
- Liao, X. (2023). Research and enlightenment on the Graduation Rate of American Universities Based on Panel Data. *International Conference on Management Engineering, Software Engineering and Service Sciences (ICMSS)*. IEEE. <https://doi.org/10.1109/ICMSS56787.2023.10118079>
- Lin, K. (2023). Research on Graduation Destination Prediction Algorithm Based on Students' Learning Behavior Data. *Asian Conference on Artificial Intelligence Technology (ACAIT)*. IEEE. <https://doi.org/10.1109/ACAIT60137.2023.10528473>
- Nurdin, H., Sartini, Sumarna, Maulana, Y. I., &



- Riyanto, V. (2023). Prediction of Student Graduation with the Neural Network Method Based on Particle Swarm Optimization. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, 8(4), 2353–2362. <https://doi.org/10.33395/sinkron.v8i4.12973>
- Ong, D. P., & Pedrasa, J. R. I. (2021). Student Risk Assessment: Predicting Undergraduate Student Graduation Probability Using Logistic Regression, SVM, and ANN. *TENCON 2021 - 2021 IEEE Region 10 Conference (TENCON)*. IEEE. <https://doi.org/10.1109/TENCON54134.2021.9707322>
- Pangesti, W. E., Ariyati, I., Priyono, Sugiono, & Suryadithia, R. (2024). Utilizing Genetic Algorithms To Enhance Student Graduation Prediction With Neural Networks. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, 9(1), 276–284. <https://doi.org/https://doi.org/10.33395/sinkron.v9i1.13161> e-ISSN
- Rawatlal, R., Chetty, M., & Naicker, A. K. (2022). Latent Factors for Consistently Predicting Student Success. *World Engineering Education Forum - Global Engineering Deans Council (WEEF-GEDC)*. IEEE. <https://doi.org/10.1109/WEEF-GEDC54384.2022.9996209>
- Ridwansyah, R., Wijaya, G., & Purnama, J. J. (2020). Hybrid Optimization Method Based on Genetic Algorithm for Graduates Students. *Jurnal Pilar Nusa Mandiri*, 16(1), 53–58. <https://doi.org/10.33480/pilar.v16i1.1180>
- Riyanto, V., Hamid, A., & Ridwansyah. (2019). Prediction of Student Graduation Time Using the Best Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(2), 1–9. <https://doi.org/http://dx.doi.org/10.24014/ijaidm.v2i1.6424>
- Rukiastindari, S., Rohimah, L., Aprillia, A., & Mutia, F. (2024). Predicting Graduation Outcomes: Decision Tree Model Enhanced with Genetic Algorithm. *Paradigma - Jurnal Komputer Dan Informatika*, 26(1), 1–6. <https://doi.org/10.31294/p.v26i1.3165>
- Salam, A., Zeniarja, J., & Anthareza, D. M. (2022). Student Graduation Prediction Model using Deep Learning Convolutional Neural Network (CNN). *International Seminar on Application for Technology of Information and Communication (ISemantic)*. IEEE. <https://doi.org/10.1109/iSemantic55962.2022.9920449>
- Suhardjono, Wijaya, G., & Hamid, A. (2019). PREDIKSI WAKTU KELULUSAN MAHASISWA MENGGUNAKAN SVM BERBASIS PSO. *Bianglala Informatika*, 7(2), 97–101. <https://doi.org/https://doi.org/10.31294/bi.v7i2.6654.g3731>
- Yennimar, Leonardi, W., Weide, H., Cantona, D., & Gani Mores Hutagalung. (2024). Comparison of data mining algorithms (random forest, C4.5, catboost) based on adaptive boosting in predicting diabetes mellitus. *Jurnal Teknik Informatika C.I.T Medicom*, 16(3), 01–12. Retrieved from [www.medikom.iocspublisher.orgjournalhomepage:www.medikom.iocspublisher.org](http://www.medikom.iocspublisher.orgjournalhomepage:www.medikom.iocspublisher.org)