

Algoritma *Naive Bayes* untuk Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Politeknik TEDC Bandung

Nandhita Ayusari^{1*}, Castaka Agus Sugianto²

^{1,2}Program Studi Teknik Informatika, Politeknik TEDC

^{1,2}Jln. Pesantren-Cibabat, Kota Cimahi, 40513, Indonesia

E-mail: nandhitaayusari002@gmail.com¹, castaka@poltektedc.ac.id²

Abstrak

Info Naskah:

Naskah masuk:

13 Desember 2024

Direvisi:

6 Maret 2025

Diterima:

7 Maret 2025

Politeknik TEDC Bandung adalah sebuah institusi pendidikan tinggi yang telah berkomitmen untuk meningkatkan efisiensi dan efektivitas dalam proses pendidikan dan pengajaran dengan mengimplementasikan berbagai kebijakan dan program. Hal ini salah satunya bertujuan untuk membuat lulusan mahasiswa yang berkualitas dan bermanfaat bagi masyarakat. Salah satu langkah yang perlu ditempuh mahasiswa agar menjadi lulusan yang berkualitas adalah dengan lulus tepat waktu. Namun pada pelaksanaannya, tetap saja ada beberapa mahasiswa yang mengalami kendala untuk mencapai hal tersebut. Hal ini karena disebabkan oleh beberapa faktor, sehingga perlu adanya upaya untuk mengurangi bahkan mengatasi hal tersebut. Penelitian ini bertujuan menerapkan algoritma *Naive Bayes* guna dapat mengklasifikasikan data mahasiswa yang tepat waktu dan tidak tepat waktu dalam mengikuti sidang tugas akhir, agar dapat memperoleh solusi dan upaya yang dapat membantu pihak kampus untuk mengatasi masalah tersebut. Hasil pengujian dengan metode *naïve bayes* tanpa validasi menghasilkan 218 data yang termasuk dalam class tepat waktu dan 33 data yang termasuk dalam class tidak tepat waktu. Sedangkan hasil hasil pengujian dengan metode *naïve bayes* menggunakan validasi menghasilkan 216 data yang termasuk dalam class tepat waktu dan 35 data yang termasuk dalam class tidak tepat waktu. Hasil pengujian dengan metode *naïve bayes* menggunakan validasi dengan *RapidMiner* menghasilkan tingkat akurasi (*accuracy*) 92,05%, *precision* memiliki nilai 92,40%, *recall* memiliki nilai 98,52% dan *F1-Score* memiliki nilai 95,71%.

Abstract

Keywords:

Polytechnic TEDC Bandung;

Timeliness;

On-time;

Naive Bayes.

Polytechnic TEDC Bandung is a higher education institution that is committed to increasing efficiency and effectiveness in the education and teaching process by implementing various policies and programs. One of the aims of this is to produce quality student graduates who are useful for society. One of the steps that students need to take to become quality graduates is to graduate on time. However, in its implementation, there are still some students who experience obstacles in achieving this. This is caused by several factors, so efforts are needed to reduce or even overcome this. This research aims to apply the Naive Bayes algorithm to be able to classify student data that is on time and not on time when attending the final assignment session, in order to obtain solutions and efforts that can help the campus to overcome this problem. The test results using the naïve Bayes method without validation produced 218 data that were included in the on-time class and 33 data that were included in the not-on-time class. Meanwhile, the results of testing using the naïve Bayes method using validation produced 216 data that were included in the on-time class and 35 data that were included in the not-on-time class. Test results using the Naïve Bayes method using validation with RapidMiner produced an accuracy level of 92.05%, precision had a value of 92.40%, Recall had a value of 98.52% and F1-Score had a value of 95.71%.

*Penulis korespondensi:

Nandhita Ayusari

E-mail: nandhitaayusari002@gmail.com

1. Pendahuluan

Kemajuan teknologi informasi sudah semakin berkembang pesat disegala bidang kehidupan. Banyak sekali data yang dihasilkan oleh teknologi informasi yang canggih, mulai dari bidang industri, ekonomi, ilmu dan teknologi serta berbagai bidang kehidupan lainnya. Penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data yang berlimpah mengenai siswa dan proses pembelajaran yang dihasilkan. Institusi pendidikan perguruan tinggi, data mahasiswa dan data jumlah kelulusan mahasiswa dapat menghasilkan informasi yang berlimpah berupa jumlah kelulusan setiap tahunnya, profil, dan hasil akademik mahasiswa selama menempuh proses kegiatan belajar mengajar di perguruan tinggi. Hal ini akan terjadi secara berulang pada sebuah perguruan tinggi [1].

Pendidikan tinggi memainkan peran krusial dalam membentuk sumber daya manusia yang berkualitas untuk mendukung pembangunan dan kemajuan suatu negara. Oleh karena itu, penting untuk memastikan tingkat kelulusan mahasiswa yang optimal guna mencapai tujuan tersebut. Untuk mencapai misi tersebut, Politeknik TEDC Bandung sebagai institusi pendidikan tinggi telah berkomitmen untuk meningkatkan efisiensi dan efektivitas dalam proses pendidikan dan pengajaran dengan mengimplementasikan berbagai kebijakan dan program.

Selama beberapa tahun terakhir, Politeknik TEDC Bandung telah menghadapi tantangan akademik yang cukup serius terkait ketidak tepatan waktu mahasiswa dalam menyelesaikan sidang tugas akhir. Masalah ini tidak hanya berdampak pada mahasiswa secara individu, tetapi juga mempengaruhi keseluruhan kinerja akademik institusi. Banyak mahasiswa tingkat akhir yang dihadapkan pada berbagai hambatan ketika hendak memasuki tahap sidang tugas akhir. Hambatan-hambatan ini dapat berupa masalah akademik seperti nilai yang tidak mencukupi, kesulitan dalam menyelesaikan proyek tugas akhir, hingga persoalan administratif atau kurangnya bimbingan yang memadai.

Fenomena ini mengindikasikan adanya kebutuhan yang mendesak untuk memperbaiki sistem pendampingan dan evaluasi akademik di Politeknik TEDC Bandung. Ketepatan waktu dalam menyelesaikan tugas akhir merupakan indikator penting dari keberhasilan proses pendidikan, karena hal ini tidak hanya mencerminkan kesiapan mahasiswa dalam menghadapi tantangan akademik, tetapi juga kesiapan mereka untuk terjun ke dunia kerja atau melanjutkan pendidikan ke jenjang yang lebih tinggi.

Penelitian ini memfokuskan pada ketepatan waktu mahasiswa untuk lulus menjadi sangat penting. Dengan mengenali faktor-faktor yang mempengaruhi ketepatan waktu lulus, Politeknik TEDC Bandung dapat merancang strategi yang lebih efektif untuk membantu mahasiswa menyelesaikan studi mereka dalam waktu yang ditentukan. Hal ini tidak hanya akan meningkatkan tingkat kelulusan, tetapi juga mempersiapkan mahasiswa dengan lebih baik untuk memasuki dunia kerja atau melanjutkan studi lanjut dengan lebih percaya diri.

Agar lebih mudah untuk mengetahui tingkat kelulusan mahasiswa tersebut dapat dilakukan dengan menggunakan teknik data *mining*. Data *mining* adalah proses interaktif untuk menemukan pola data dan memprediksi perilaku masa depan berdasarkan pola yang ada. Pada dasarnya, data *mining* melibatkan analisis data dan penggunaan teknik perangkat lunak untuk mencari pola dalam kumpulan data tersembunyi. Dengan memperoleh data yang ada, hubungan antar *item* maupun faktor-faktor tertentu dalam penyelenggaraan proses belajar mengajar dapat digali atau dianalisis dan diselidiki lebih lanjut [2].

Beberapa teknik data mining telah dikembangkan dan diimplementasikan untuk mengekstrak pengetahuan dan informasi untuk menemukan pola pengetahuan yang mungkin berguna untuk pengambilan keputusan. Teknik-teknik yang digunakan untuk pengekstrakan pengetahuan dalam data *mining* adalah pengenalan pola, *clustering*, asosiasi, prediksi dan klasifikasi. Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu [3].

Salah satu algoritma yang dapat digunakan untuk klasifikasi adalah *Naive Bayes*. Algoritma *Naive Bayes* adalah teknik pengelompokan yang menggunakan prinsip probabilitas dan statistik. Pengelompokan dengan *Naïve Bayes* adalah metode statistik yang mampu memperkirakan kemungkinan keanggotaan dalam suatu kategori. Metode ini dikenal memiliki tingkat akurasi yang tinggi dan dapat beroperasi dengan cepat saat digunakan pada kumpulan data yang besar [4]. Algoritma menggunakan teorema *Bayes* dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas [5]. Kelebihan dari algoritma *Naive Bayes* terletak pada kesederhanaan dan kecepatannya yaitu dengan struktur yang relatif sederhana, algoritma ini memungkinkan pelatihan dan prediksi yang cepat, ideal untuk dataset besar atau saat sumber daya komputasi terbatas [6].

Berdasarkan penelitian yang dilakukan oleh Castaka Agus Sugianto pada tahun 2019, kelebihan algoritma *Naive Bayes* dibandingkan dengan *Decision Tree* adalah bahwa *Naive Bayes* dapat bekerja dengan cepat dan efisien ketika diterapkan pada dataset yang besar. Selain itu, meskipun *Decision Tree* memiliki tingkat akurasi yang lebih tinggi dalam contoh yang diberikan (73,97% dibandingkan dengan 58,29% untuk *Naive Bayes*), *Naive Bayes* juga terbukti memiliki tingkat akurasi yang tinggi. Kecepatan dan efisiensi *Naive Bayes* dalam memproses data besar menjadikannya unggul dalam situasi tertentu di mana pemrosesan cepat dan akurasi cukup tinggi diperlukan [7].

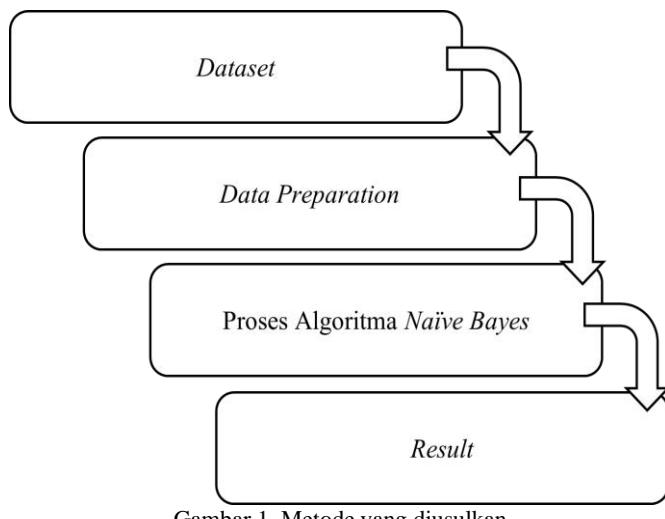
Algoritma *Naive Bayes* adalah salah satu algoritma terbaik jika dibandingkan dengan algoritma lain seperti algoritma *logistic regression*, *neural network*, *support vector machine* dan *k-nearest neighbour*. Hal ini sebab algoritma *Naïve Bayes* memiliki efisiensi dalam menangani banyak fitur, membuatnya cocok untuk *dataset* yang kompleks [8]. *Naive Bayes* juga efisien dalam mengelola data kategorikal, yang sering ditemukan dalam prediksi kelulusan mahasiswa, seperti

jenis kelamin, jurusan, atau kategori nilai. Meskipun asumsi independensi fitur, algoritma ini tetap memberikan hasil yang baik, terutama ketika asumsi tersebut tidak sepenuhnya dilanggar. Kelebihan lainnya mencakup kemampuannya untuk bekerja dengan *dataset* kecil, interpretasinya yang mudah, dan ketangguhannya terhadap *noise* dalam data. Meski begitu, perlu diingat bahwa asumsi independensi fitur dapat menjadi kendala, dan performanya bisa kalah jika asumsi tersebut tidak terpenuhi. Meskipun demikian, *Naïve Bayes* tetap menjadi pilihan yang baik untuk tugas klasifikasi seperti prediksi kelulusan mahasiswa [9].

Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi kinerja algoritma *Naïve Bayes* dalam prediksi kelulusan mahasiswa yaitu untuk mengetahui sejauh mana keunggulan algoritma *Naïve Bayes* dalam memproses dan menganalisis data mahasiswa, dengan fokus pada prediksi kelulusan. Metode yang digunakan adalah *cross validation* yaitu metode evaluasi yang digunakan untuk menilai kinerja model dengan lebih akurat dan mengurangi kemungkinan *overfitting*. *Cross validation*, data dibagi menjadi beberapa *subset* atau "*folds*". Pada setiap iterasi, satu *fold* digunakan sebagai data uji, sementara sisa *fold* lainnya digunakan untuk melatih model. Proses ini diulang hingga setiap *fold* telah digunakan sekali sebagai data uji. Hasil evaluasi dari setiap iterasi kemudian dirata-rata untuk memberikan estimasi kinerja model yang lebih stabil dan dapat diandalkan. *Cross validation* membantu memastikan bahwa model *Naïve Bayes* memiliki kinerja yang konsisten dan generalisasi yang baik ketika diterapkan pada data baru yang tidak terlihat selama pelatihan sehingga dapat memperoleh tingkat keakuratan data yang lebih baik [10].

2. Metode

Metode yang digunakan dalam pengolahan data dengan algoritma *Naïve Bayes* pada penelitian ini, dapat dilihat gambar 1.



Gambar 1. Metode yang diusulkan

Berdasarkan gambar 1, terdapat beberapa tahapan yaitu diantaranya sebagai berikut.

2.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah data terkait keikutsertaan mahasiswa Politeknik TEDC Bandung dalam mengikuti sidang tugas akhir yaitu data kelulusan tahun 2019 - 2023 dan data dari kuisioner yang disebarluaskan secara *online* kepada alumni, mulai dari mahasiswa yang tepat waktu hingga yang terlambat untuk mengikuti sidang tugas akhir. Data yang diperoleh adalah 251 data, dengan variabel yang terdiri atas no, nama lengkap, jenis kelamin, program studi, angkatan, hobi, pekerjaan ayah, pekerjaan ibu, pendapatan ayah, pendapatan ibu, uang saku per bulan, tempat tinggal, status kuliah sambil kerja, status kuliah cuti, status masalah keuangan, status organisasi, jumlah organisasi, status nilai bermasalah, jumlah nilai bermasalah, total bimbingan, tanggal sidang dan ipk. Agar dapat mengetahuinya lebih jelas, dapat melihat pada tabel 1 dibawah ini.

Tabel 1. Data Sidang Tugas Akhir Mahasiswa

No.	Nama Lengkap	Jenis Kelamin	...	IPK
1.	Shintya Putri Rahayu R.A. Melia Sri	Perempuan	...	3.73
2.	Aryanti	Perempuan	...	3.61
3.	Nurul Ulfa Hasanah	Perempuan	...	3.62
4.	Tiara Sekar Diva Chatriya Maulana	Perempuan	...	3.59
5.	Ibrahim	Laki-laki	...	3.65
6.	Halimatus Sadiyah	Perempuan	...	3.73
7.	Dwi Novi Hermarini	Perempuan	...	3.60
8.	Silvia Suciawati	Perempuan	...	3.48
:	:	:	:	:
251	Muchammad Miftach Machyudiono	Laki-Laki	...	3.37

2.2 Data Preparation

Preparasi data adalah proses persiapan data sebelum data tersebut digunakan untuk analisis. Tujuan utama dari preparasi data adalah untuk memastikan bahwa data yang digunakan adalah berkualitas tinggi, relevan, dan siap digunakan untuk tujuan tertentu. Jumlah dataset pada penelitian ini adalah 251 *record* dan variabel yang digunakan berjumlah 22 yang terdiri atas no, nama lengkap, jenis kelamin, program studi, angkatan, hobi, pekerjaan ayah, pekerjaan ibu, pendapatan ayah, pendapatan ibu, uang saku per bulan, tempat tinggal, status kuliah sambil kerja, status kuliah cuti, status masalah keuangan, status organisasi, jumlah organisasi, status nilai bermasalah, jumlah nilai bermasalah, total bimbingan, tanggal sidang dan ipk.

Penulis menyesuaikan agar variabel yang akan digunakan hanya terdiri atas 14 variabel yaitu pendapatan ayah, pendapatan ibu, uang saku per bulan, tempat tinggal, status kuliah sambil kerja, status kuliah cuti, status masalah keuangan, status organisasi, jumlah organisasi, status nilai bermasalah, jumlah nilai bermasalah, total bimbingan, tanggal sidang dan ipk. Data yang telah dipreparasi dapat dilihat pada tabel 2.

Tabel 2. Hasil Preparasi Data

No	Pendapatan Ayah	Pendapatan Ibu	Uang Saku Per Bulan	Tempat Tinggal	Status Kuliah Sambil Kerja	Status Kuliah Cuti	:	IPK
1	4	4	800.000	0	Ya	Tidak	...	3.73
2	2	2	400.000	1	Tidak	Tidak	...	3.61
3	2	2	600.000	0	Ya	Tidak	...	3.62
4	3	4	700.000	0	Tidak	Tidak	...	3.59
5	3	2	450.000	1	Ya	Tidak	...	3.65
6	2	4	550.000	0	Tidak	Tidak	...	3.73
7	2	2	650.000	1	Ya	Tidak	...	3.60
:	:	:	:	:	:		:	:
251	3	0	900.000	1	Tidak	Tidak	...	3.37

2.3 Proses Algoritma *Naïve Bayes*

Metode *Naïve Bayes* adalah salah satu model klasifikasi probalistik sederhana guna menghitung kumpulan probabilitas dan menjumlahkan frekuensi dan kombinasi dari *dataset*. *Naïve Bayes* mempunyai beberapa keuntungan antara lain hanya membutuhkan jumlah *training* data yang kecil. Training data digunakan menentukan estimasi parameter dalam proses klasifikasi [11].

Adapun persamaan teorema metode *Naïve Bayes* adalah [11]:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Di mana:

X : *Class* data yang belum diketahui.

H : Spesifik *class* yang merupakan hipotesis data

$P(H|X)$: Posteriori probabilitas (probabilitas hipotesis H berdasarkan kondisi X)

$P(H)$: Berisi nilai probabilitas hipotesis H

$P(X|H)$: Berisi nilai probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Berisi nilai probabilitas hipotesis X

Selain itu perhitungan juga dilakukan dengan menggunakan rumus *F1-Score* untuk dapat mengetahui nilai *accuracy*, *precision*, dan *recall* yaitu [12]:

$$1) \quad \text{Accuracy} \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2) Precision

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

3) Recall

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

4) F1-Score

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Keterangan :

True Positives (TP): Jumlah *instance* yang benar-benar positif dan diprediksi positif.

False Positives (FP): Jumlah *instance* yang sebenarnya negatif tetapi diprediksi positif.

True Negatives (TN): Jumlah *instance* yang benar-benar negatif dan diprediksi negatif.

False Negatives (FN): Jumlah *instance* yang sebenarnya positif tetapi diprediksi negatif.

2.4 Result

Hasil dari penelitian ini adalah klasifikasi dari data mahasiswa terkait keterlambatan mengikuti sidang tugas akhir di Politeknik TEDC. Dalam penelitian ini juga akan dihasilkan pengujian algoritma *Naïve Bayes* dengan *cross validation* pada data mahasiswa terkait keterlambatan mengikuti sidang tugas akhir di Politeknik TEDC dan memperoleh kesimpulan dari tiap-tiap atribut yang digunakan pada pengujian.

3. Hasil dan Pembahasan

Adapun hasil dan pembahasan pada penelitian yang telah dilakukan akan dipaparkan sebagai berikut.

3.1 Hasil Penelitian

Hasil penelitian membahas proses pengolahan data awal, pengujian algoritma *naïve bayes* untuk melihat *class* tingkat ketepatan waktu kelulusan mahasiswa dan pengujian menggunakan metode *cross validation*.

1) Proses Pengolahan Data Awal

Proses pengelolaan data awal pada pengujian ini menggunakan perangkat lunak *RapidMiner*. Tahapan ini dilakukan dengan memasukan data yang sudah dipreparasi sebelumnya dengan cara *import* data pada perangkat lunak *RapidMiner*. Dalam tahap ini, data mentah yang telah disiapkan sebelumnya diolah dan dimasukkan ke dalam *RapidMiner* untuk dianalisis lebih lanjut. Data yang diimport berasal dari sumber “.xlsx”. Selanjutnya, data yang sudah

diimport akan melalui berbagai tahap pengolahan seperti pembersihan data, transformasi data, dan penyesuaian format agar siap digunakan dalam proses analisis dan pemodelan selanjutnya.

Pada tahap penyesuaian format, penulis akan menyesuaikan format dari setiap kolom data yang akan diolah yaitu pendapatan ayah memiliki format *polynomial*, pendapatan ibu memiliki format *polynomial*, uang saku per bulan memiliki format *integer*, tempat tinggal memiliki format *binomial*, status kuliah sambil kerja memiliki format *binomial*, status kuliah cuti memiliki format *binomial*, status masalah keuangan memiliki format *binomial*, status organisasi memiliki format *binomial*, jumlah organisasi memiliki format *integer*, status nilai bermasalah memiliki format *binomial*, jumlah nilai bermasalah memiliki format *integer*, total bimbingan memiliki format *integer*, tanggal sidang memiliki format *binomial*, dan ipk memiliki format *real*. Adapun data preparasi yang telah diimport pada *RapidMiner* dapat dilihat pada gambar 2 berikut.

Gambar 2. Data Preparasi yang Telah Diimport

2) Pengujian Algoritma *Naive Bayes* Tanpa Validasi

Setelah proses tersebut penulis membutuhkan pengujian data menggunakan perangkat lunak *RapidMiner* untuk melihat *class performance* tingkat ketepatan waktu menggunakan algoritma *naive bayes* dengan memasukan data testing agar data tersebut dapat digunakan sebagai data klasifikasi dengan ketentuan data yang digunakan harus sama *attribute* dengan data *training*, baik dari urutan maupun jenis datanya dengan proses yang sama pada seperti data *training*. Selanjutnya mengimport data testing tadi ke dalam *RapidMiner*. Kemudian masukkan data *testing* tersebut ke dalam halaman *process*. Lalu cari *apply model* di operator dan tarik ke halaman *process* setelah itu cari *performance* di operator dan tarik ke halaman *process*. Kemudian *naive bayes* sambungkan ke *apply model*, data testing sambungkan ke *apply model*, *apply model* sambungkan ke *performance* dan *performance* sambungkan ke *result*.

Selanjutnya menekan tombol *Run Process* untuk memulai proses pengujinya. Hasil pengujian data dengan algoritma *Naive Bayes* dari 251 data mahasiswa adalah 218 data yang termasuk *class* tepat waktu dan 33 data yang termasuk *class* tidak tepat waktu.

Untuk hasil *performance* dari pengujian data yang dilakukan memiliki nilai akurasi (*accuracy*) sebesar 92,83%. Untuk lebih jelasnya dapat dilihat pada gambar 3 berikut.

Gambar 3. Hasil Accuracy Tanpa Validasi

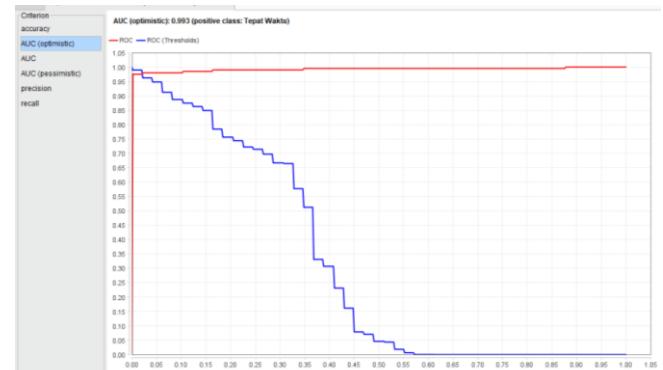
Adapun hasil *precision* pada pengujian ini adalah sebesar 92,2%. Hasilnya dapat dilihat pada gambar 4 berikut.

Gambar. 4 Hasil Precision Tanpa Validasi

Hasil dari *Recall* pada pengujian ini adalah sebesar 99,5%. Hasilnya dapat dilihat pada gambar 5 berikut.

Gambar. 5 Hasil Recall Tanpa Validasi

Berdasarkan hasil pengujian tanpa validasi, tingkat akurasi AUC memiliki nilai sebesar 0,993. Hasil tersebut menunjukkan hasil pengujian ini termasuk kedalam kategori *Excellent Classification* (Sangat Baik). Adapun hasil AUC dapat dilihat pada gambar 6 berikut.



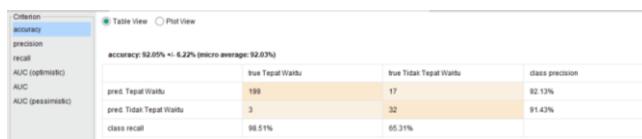
Gambar. 6 Hasil AUC Tanpa Validasi

3) Pengujian Algoritma *Naive Bayes* Menggunakan Metode Cross Validation

Pengujian algoritma *Naive Bayes* menggunakan cross-validation membagi dataset menjadi beberapa lipatan (folds). Data dibagi acak ke dalam k lipatan, algoritma dilatih pada k-1 lipatan dan diuji pada lipatan sisanya. Proses diulang 10 kali

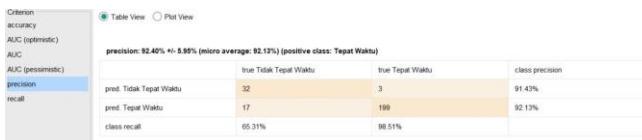
sehingga setiap lipatan digunakan sekali sebagai data uji. Rata-rata hasil pengujian memberikan estimasi performa yang lebih akurat dan mengurangi bias. Metode ini memastikan model memiliki kemampuan generalisasi yang baik pada data baru.

Untuk hasil *performance* dari pengujian data yang dilakukan menggunakan metode *cross validation* dengan jumlah lipatan (*folds*) 10 (*10-fold cross validation*) yaitu dari 251 data mahasiswa terdapat 216 data yang termasuk *class* tepat waktu dan 35 data yang termasuk *class* tidak tepat waktu dengan tingkat akurasi 92,05%. Untuk lebih jelasnya dapat dilihat pada gambar 7 berikut.



Gambar. 7 Hasil Accuracy Metode Cross Validation

Adapun hasil *precision* pada pengujian ini adalah sebesar 92,13%. Hasilnya dapat dilihat pada gambar 8 berikut.



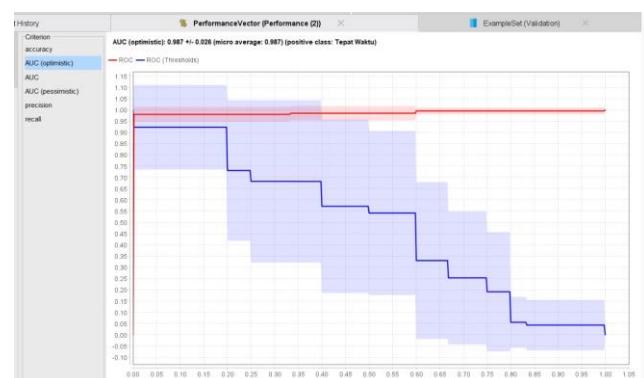
Gambar. 8 Hasil Precision Metode Cross Validation

Hasil dari *Recall* pada pengujian ini adalah sebesar 98,51%. Hasilnya dapat dilihat pada gambar 9 berikut.



Gambar. 9 Hasil Recall Metode Cross Validation

Berdasarkan hasil pengujian dengan metode *cross validation*, tingkat akurasi AUC memiliki nilai sebesar 0,987. Hasil tersebut menunjukkan hasil pengujian ini termasuk kedalam kategori *Excellent Classification* (Sangat Baik). Adapun hasil AUC dapat dilihat pada gambar 10 berikut.



Gambar. 10 Hasil AUC Metode Cross Validation

3.2 Pembahasan

Pada bagian pembahasan akan menjelaskan terkait analisis perhitungan *performance* secara manual dan analisis klasifikasi algoritma *naïve bayes* dengan menggunakan aplikasi serta kesimpulannya. Hal ini dilakukan untuk mengetahui kesesuaian hasil klasifikasi sehingga menghasilkan data akurat untuk klasifikasi data mahasiswa yang tepat waktu dan tidak tepat waktu.

1) Analisis Perhitungan Performance Secara Manual

Perhitungan *performance* secara manual dilakukan dengan menggunakan rumus *F1-Score*. *F1-Score* merupakan metrik yang menggabungkan *precision* dan *recall* menjadi satu nilai yang harmonis. Metode ini sangat penting dalam evaluasi model klasifikasi, terutama ketika terdapat ketidakseimbangan kelas atau ketika keseimbangan antara *precision* dan *recall* sangat krusial. Dari 251 data mahasiswa terdapat 202 data yang termasuk dalam kelas tepat waktu dan 49 data yang termasuk dalam kelas tidak tepat waktu.

Adapun tabel *confusion matrix* dari 251 data mahasiswa untuk klasifikasi tepat waktu dan tidak tepat waktu tanpa validasi dapat dilihat pada tabel 3 berikut.

Tabel 3. Confusion Matrix Tanpa Validasi

Actual	Prediksi	
	Tepat Waktu (Positive)	Tidak Tepat Waktu (Negative)
Tepat Waktu	201 (TP)	17 (FP)
Tidak Tepat Waktu	1 (FN)	32 (TN)

Berdasarkan tabel 3 maka *accuracy*, *precision*, *recall* dan *F1-Score* dapat dihitung dan diperoleh hasil perhitungan tanpa validasi sebagai berikut.

• Precision

$$\begin{aligned} \text{Precision} &= \frac{201}{201+17} \times 100\% \\ &= 0,922 \times 100\% \\ &= 92,2\% \end{aligned}$$

• Recall

$$\begin{aligned} \text{Recall} &= \frac{201}{201+1} \times 100\% \\ &= 0,995 \times 100\% \\ &= 99,5\% \end{aligned}$$

- *F1-Score*

$$\begin{aligned}
 F1\text{-Score} &= 2 \times \frac{0,922 \times 0,995}{0,922+0,995} \times 100\% \\
 &= 2 \times \frac{0,91739}{1,917} \times 100\% \\
 &= 0,9571 \times 100\% \\
 &= 95,71\%
 \end{aligned}$$

- *Accuracy*

$$\begin{aligned}
 Accuracy &= \frac{201+32}{201+32+1+17} \times 100\% \\
 &= 0,9283 \times 100\% \\
 &= 92,83\%
 \end{aligned}$$

Berdasarkan hasil perhitungan dengan menggunakan rumus *F1-Score* maka diperoleh hasil perhitungan tanpa validasi dari nilai *accuracy* yaitu 0,9283 (92,83%), *precision* yaitu 0,922 (92,2%) dan *recall* yaitu 0,995 (99,5%).

Kemudian untuk tabel *confusion matrix* dari 251 data mahasiswa untuk klasifikasi tepat waktu dan tidak tepat waktu menggunakan validasi (*cross validation*) dapat dilihat pada tabel 4 berikut.

Tabel 4. *Confusion Matrix* Menggunakan Validasi (*Cross Validation*)

Actual	Prediksi	
	Tepat Waktu (Positive)	Tidak Tepat Waktu (Negative)
Tepat Waktu	199 (TP)	17 (FP)
Tidak Tepat Waktu	3 (FN)	32 (TN)

Berdasarkan tabel 4 maka *accuracy*, *precision*, *recall* dan *F1-Score* dapat dihitung dan diperoleh hasil perhitungan dengan validasi (*cross validation*) sebagai berikut.

- *Precision*

$$\begin{aligned}
 Precision &= \frac{199}{199+17} \times 100\% \\
 &= 0,9213 \times 100\% \\
 &= 92,13\%
 \end{aligned}$$

- *Recall*

$$\begin{aligned}
 Recall &= \frac{199}{199+3} \times 100\% \\
 &= 0,9851 \times 100\% \\
 &= 98,51\%
 \end{aligned}$$

- *F1-Score*

$$\begin{aligned}
 F1\text{-Score} &= 2 \times \frac{0,9213 \times 0,9851}{0,9213+0,9851} \times 100\% \\
 &= 2 \times \frac{0,9076}{1,9064} \times 100\% \\
 &= 0,9521 \times 100\% \\
 &= 95,21\%
 \end{aligned}$$

- *Accuracy*

$$\begin{aligned}
 Accuracy &= \frac{199+32}{199+32+3+17} \times 100\% \\
 &= 0,9205 \times 100\% \\
 &= 92,05\%
 \end{aligned}$$

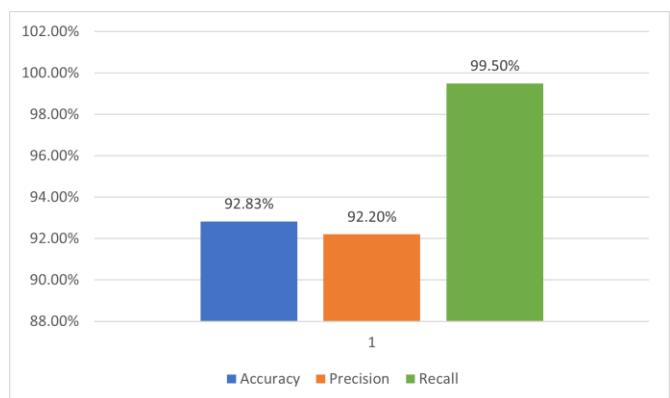
Berdasarkan hasil perhitungan dengan menggunakan rumus *F1-Score* maka diperoleh hasil perhitungan dengan validasi (*cross validation*) dari nilai *accuracy* yaitu 0,9205

(92,05%), *precision* yaitu 0,9213 (92,13%) dan *recall* yaitu 0,9851 (98,51%).

2) Analisis Klasifikasi Algoritma *Naïve Bayes* Menggunakan Aplikasi

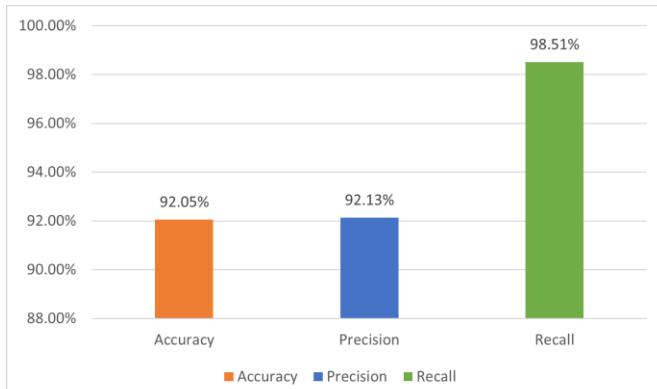
Berdasarkan data awal sebelum pengujian, dari 251 data mahasiswa terdapat 202 data yang termasuk dalam kelas tepat waktu dan 49 data yang termasuk dalam kelas tidak tepat waktu. Setelah dilakukan pengujian dengan *software RapidMiner* menggunakan 14 atribut, yaitu pendapatan ayah, pendapatan ibu, uang saku per bulan, tempat tinggal, status kuliah sambil kerja, status cuti kuliah, status masalah keuangan, status organisasi, jumlah organisasi, status nilai bermasalah, jumlah nilai bermasalah, total bimbingan, tanggal sidang, dan IPK, hasil klasifikasi menjadi lebih baik dan akurat.

Hasil pengujian algoritma *naïve bayes* tanpa validasi menghasilkan 218 data yang termasuk dalam *class* tepat waktu dan 33 data yang termasuk dalam *class* tidak tepat waktu. Hasil pengujian dengan metode *naïve bayes* memiliki *accuracy* yaitu 92,83%, dengan *precision* sebesar 92,2% dan *recall* sebesar 99,5%. Untuk lebih jelasnya dapat dilihat pada gambar 11 berikut.



Gambar. 11 Nilai *Accuracy*, *Precision* dan *Recall* dengan Aplikasi Tanpa Validasi

Hasil pengujian algoritma *naïve bayes* dengan validasi melalui *software RapidMiner* yaitu dengan menggunakan *cross validation* (*10-fold cross validation*) dari 251 data menghasilkan 216 data yang termasuk dalam *class* tepat waktu dan 35 data yang termasuk dalam *class* tidak tepat waktu dengan tingkat akurasi 92,05%, *precision* memiliki nilai 92,13% dan *Recall* memiliki nilai 98,51%. Untuk lebih jelasnya dapat dilihat pada gambar 12 berikut.



Gambar. 12 Nilai *Accuracy*, *Precision* dan *Recall* dengan Aplikasi Menggunakan Validasi

3) Kesimpulan Klasifikasi Algoritma *Naïve Bayes*

Dari 251 data mahasiswa, setelah dilakukan perhitungan secara manual hingga pengujian menggunakan aplikasi dengan menggunakan 14 atribut, yaitu pendapatan ayah, pendapatan ibu, uang saku per bulan, tempat tinggal, status kuliah sambil kerja, status cuti kuliah, status masalah keuangan, status organisasi, jumlah organisasi, status nilai bermasalah, jumlah nilai bermasalah, total bimbingan, tanggal sidang, dan IPK menghasilkan nilai *accuracy*, *precision* dan *recall* yang tidak jauh berbeda.

Adapun untuk hasil klasifikasi nilai *accuracy*, *precision* dan *recall* dari perhitungan *performance* secara manual, dengan menggunakan aplikasi tanpa validasi dan menggunakan validasi dapat dilihat pada tabel 5 berikut.

Tabel 5. Hasil Klasifikasi

Keterangan	Klasifikasi	
	Dengan Aplikasi Tanpa Validasi	Dengan Aplikasi Menggunakan Validasi
Tepat Waktu	218	216
Tidak Tepat Waktu	33	35

Berdasarkan gambar 10 maka klasifikasi menunjukkan hasil yang serupa, dimana kuantitas dan kualitas datanya tidak jauh berbeda. Selain itu, hasil pengujiannya juga termasuk dalam kategori *Excellent Classification* (Sangat Baik), karena memiliki nilai AUC sebesar 0,993 tanpa *cross validation* dan 0,987 menggunakan *cross validation*.

Namun perlu diketahui, hal ini dipengaruhi oleh pemilihan atribut yang digunakan. Dari 11 atribut yang digunakan dalam pengujian tersebut, saling mempengaruhi hasil pengujian, sehingga sangat penting untuk memilih dan menyesuaikan atribut yang digunakan agar memperoleh hasil klasifikasi dengan tingkat akurasi yang lebih baik.

Pemilihan atribut dalam suatu pengujian sangatlah penting karena setiap atribut dapat mempengaruhi hasil akhir dari klasifikasi. Berikut penjelasan mengenai atribut yang dipilih dalam pengujian ini:

- Pendapatan Ayah dan Pendapatan Ibu: Atribut ini dipilih karena pendapatan orang tua dapat mempengaruhi kemampuan finansial mahasiswa dalam menyelesaikan studi tepat waktu.
- Uang Saku per Bulan: Besaran uang saku bisa berpengaruh terhadap kemampuan mahasiswa dalam mengelola kebutuhan sehari-hari yang mendukung kegiatan akademik mereka.
- Tempat Tinggal: Faktor tempat tinggal, seperti tinggal bersama orang tua atau kos, dapat mempengaruhi fokus dan konsentrasi mahasiswa dalam belajar.
- Status Kuliah Sambil Kerja: Mahasiswa yang bekerja sambil kuliah mungkin memiliki waktu yang lebih sedikit untuk belajar, sehingga mempengaruhi kelulusan tepat waktu.
- Status Cuti Kuliah: Mahasiswa yang pernah mengambil cuti cenderung memiliki kemungkinan yang lebih besar untuk tidak lulus tepat waktu.
- Status Masalah Keuangan: Masalah keuangan bisa menjadi hambatan dalam menyelesaikan studi tepat waktu.
- Status Organisasi dan Jumlah Organisasi: Keterlibatan dalam organisasi mahasiswa dapat mempengaruhi waktu belajar dan komitmen akademik mahasiswa.
- Status Nilai Bermasalah dan Jumlah Nilai Bermasalah: Nilai akademik yang bermasalah dapat menjadi indikator kemampuan akademik mahasiswa dan potensi kelulusan mereka.
- Total Bimbingan: Jumlah bimbingan yang diterima mahasiswa dapat mencerminkan dukungan akademik yang mereka terima, yang berpengaruh pada keberhasilan studi.
- Tanggal Sidang: Jadwal sidang tugas akhir dapat mempengaruhi kelulusan mahasiswa.
- IPK (Indeks Prestasi Kumulatif): IPK merupakan indikator utama prestasi akademik mahasiswa yang langsung berhubungan dengan kelulusan tepat waktu.

4. Kesimpulan

Berdasarkan penelitian dan pengujian, algoritma *Naïve Bayes* berhasil mengklasifikasikan ketepatan waktu kelulusan mahasiswa dengan akurat. Dari 251 data, 202 mahasiswa dikategorikan lulus tepat waktu dan 49 tidak tepat waktu menggunakan *F1-Score*. Tanpa validasi, algoritma ini menghasilkan 218 data tepat waktu dan 33 tidak tepat waktu; dengan validasi, 216 data tepat waktu dan 35 tidak tepat waktu. Hasil ini serupa dalam kualitas dan kuantitas data, serta masuk kategori *Excellent Classification* dengan AUC 0,993 tanpa *cross validation* dan 0,987 menggunakan *cross validation*. Performanya mencakup akurasi 92,83%, *precision* 92,2%, dan *recall* 99,5% tanpa validasi, serta akurasi 92,05%, *precision* 92,13%, dan *recall* 98,51% dengan validasi.

Saran Untuk pengembangan selanjutnya, bisa menggunakan metode atau algoritma lain seperti *C4.5* atau *K-Nearest Neighbor* untuk membandingkan nilai akurasi. Tambahkan atribut pendukung lain dalam pengujian data,

seperti kehadiran perkuliahan. Meningkatkan jumlah data untuk *training* model juga diharapkan dapat meningkatkan performa akurasi.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada Kemendikbud atas beasiswa yang diberikan, Dr. Gerinata Ginting, S.E., M.M. atas kebijakan pembelajaran, Castaka Agus S., M.Kom., MCS., Wakil Direktur I Bidang Akademik dan Ketua Program Studi Teknik Informatika yang juga menjadi pembimbing dalam penyusunan tugas akhir, dosen Teknik Informatika serta staf Politeknik TEDC Bandung, dan kepada orang tua tercinta, Papa Agustinus Ruruk dan Ibu Yuliana Rissing. Terima kasih juga ditujukan kepada Om Juprianto Rissing, S.H., Om Nader Rissing, Jerianton Rissing, Bunda Setriana, adik-adik Dion, Diva, Diki, serta teman dan sahabat Aristevani dan Djessica. Kritik dan saran sangat diharapkan demi kesempurnaan laporan ini.

Daftar Pustaka:

- [1] Y. S. Nugroho, "Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro," *J. Semant.* 2019, vol. 1, no. 3, pp. 1–11, 2019.
- [2] A. M. Sitompul, S. Tunas Bangsa, S. Utara, A. Tunas Bangsa, and I. A. Jln Sudirman Blok No, "Teknik Data Mining Dalam Prediksi Jumlah Siswa Baru Dengan Algoritma Naive Bayes," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 2, no. 2, pp. 108–117, 2021.
- [3] S. Defiyanti, "Integrasi Metode Klasifikasi Dan Clustering dalam Data Mining," pp. 39–44, 2015.
- [4] A. Himawan and C. A. Sugianto, "Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor (KNN) Pada Video Youtube Mengenai Global Warming," *Informatics Digit. Expert*, vol. 2, pp. 98–104, 2024.
- [5] E. Manalu, F. A. Siantur, and M. R. Manalu, "Penerapan Algoritma Naïve Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV Papadan Mama Pastries," *J. Mantik Penusa*, vol. 1, no. 2, pp. 16–21, 2017.
- [6] Y. I. Kurniawan, "Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 455–464, 2021, doi: 10.25126/jtiik.201854803.
- [7] C. A. Sugianto and F. R. Maulana, "Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama)," *Techno.Com*, vol. 18, no. 4, pp. 321–331, 2019, doi: 10.33633/tc.v18i4.2587.
- [8] T. H. Apandi and C. A. Sugianto, "Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP," *J. Inform.*, vol. 7, no. 2, pp. 125–128, 2019.
- [9] S. Syarli and A. A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2018.
- [10] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [11] W. E. Nugroho, A. Sofyan, and O. Somantri, "Metode Naive Bayes Dalam Menentukan Program Studi Bagi Calon Mahasiswa Baru," *Infotekmesin*, vol. 12, no. 1, pp. 59–64, 2021, doi: 10.35970/infotekmesin.v12i1.491.
- [12] N. Khasanah, A. Salim, N. Afni, R. Komarudin, and Y. I. Maulana, "Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes," *Technol. J. Ilm.*, vol. 13, no. 3, pp. 207–214, 2022, doi: 10.31602/tji.v13i3.7312.