

PREDIKSI KELULUSAN TEPAT WAKTU MAHASISWA DENGAN METODE NAÏVE BAYES : (Studi Kasus Fakultas Ilmu Komputer Universitas Al Asyariah Mandar)

¹Muslihan, ²Ul Khairat, ³Arwansyah

^{1,2,3}Universitas Al Asyariah Mandar.

, ¹muslihan@unasman.ac.id, ²ulkhairat@mail.unasman.ac.id, ³arwan.psky112@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk memprediksi kelulusan tepat waktu mahasiswa menggunakan metode Naïve Bayes. Sistem yang dikembangkan akan mengklasifikasikan kelulusan mahasiswa dengan mengevaluasi kinerja mereka pada tahun pertama dan kedua. Data mahasiswa yang dianalisis mencakup atribut Program Studi, gender, Indeks Prestasi Semester (IPS) Semester 1-5, Indeks Prestasi Kumulatif (IPK) Semester 5, dan usia. Data ini dikumpulkan, dibersihkan, dan dianalisis menggunakan algoritma Naïve Bayes yang diimplementasikan di platform Google Colaboratory. Hasil penelitian menunjukkan bahwa metode Naïve Bayes memiliki kemampuan yang baik dalam memprediksi kelulusan tepat waktu mahasiswa dengan tingkat akurasi keseluruhan sebesar 83%. Model ini menunjukkan nilai precision 1.00, recall 0.75, dan F1-score 0.86 untuk kelas "Tidak Lulus Tepat Waktu". Untuk kelas "Lulus Tepat Waktu", model memiliki nilai precision 0.67, recall 1.00, dan F1-score 0.80. Secara keseluruhan, rata-rata makro (macro avg) menunjukkan precision 0.83, recall 0.88, dan F1-score 0.83, sementara rata-rata tertimbang (weighted avg) menunjukkan precision 0.89, recall 0.83, dan F1-score 0.84. Temuan ini menunjukkan bahwa metode Naïve Bayes dapat dijadikan alat bantu yang efektif bagi perguruan tinggi dalam memantau dan meningkatkan kelulusan tepat waktu mahasiswa.

Kata Kunci— Data Mining, Google Colaboratory, Mahasiswa, Naïve Bayes, Prediksi Kelulusan.

ABSTRACT

The objective of this research is to predict students' timely graduation using the Naïve Bayes method. The developed system classifies student graduation by evaluating their performance in the first and second years. The student data analyzed includes attributes such as Study Program, gender, Grade Point Average (GPA) for semesters 1-5, Cumulative GPA for semester 5, and age. This data is collected, cleaned, and analyzed using the Naïve Bayes algorithm implemented on the Google Colaboratory platform. The results of the study indicate that the Naïve Bayes method has a good capability in predicting timely graduation with an overall accuracy rate of 83%. For the "Not Graduating on Time" class, the model demonstrates a precision of 1.00, recall of 0.75, and F1-score of 0.86. For the "Graduating on Time" class, the model achieves a precision of 0.67, recall of 1.00, and F1-score of 0.80. Overall, the macro average shows a precision of 0.83, recall of 0.88, and F1-score of 0.83, while the weighted average indicates a precision of 0.89, recall of 0.83, and F1-score of 0.84. These findings suggest that the Naïve Bayes method can serve as an effective tool for higher education institutions to monitor and enhance students' timely graduation rates.

Keywords— Data Mining, Google Colaboratory, Graduation Prediction, Naïve Bayes, Student.

1. PENDAHULUAN

Istilah mahasiswa mengacu pada seseorang yang secara resmi terdaftar di lembaga pendidikan tinggi, baik itu negeri maupun swasta, untuk mengejar pendidikan. [1]. Mahasiswa memiliki peran penting dalam kesuksesan sebuah perguruan tinggi, baik itu swasta maupun negeri. Di Indonesia, grade akreditasi yang dikeluarkan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) menjadi penentu kualitas suatu perguruan tinggi. Terdapat banyak aspek yang dapat dijadikan sebagai indikator kualitas perguruan tinggi. [2]. Keberhasilan mahasiswa menyelesaikan studi tepat waktu menjadi salah satu indikator mutu perguruan tinggi. [3]. Pada keadaan normal, mahasiswa tingkat strata-1 atau S1 di sebuah universitas dapat menyelesaikan studi dalam waktu 4 tahun atau 8 semester. [4]. Namun, faktanya, setiap awal tahun akademik, jumlah mahasiswa yang diterima terus meningkat, namun tidak semua mahasiswa berhasil menyelesaikan studi tepat waktu sesuai jadwal yang ditentukan., sehingga mengakibatkan penambahan data mahasiswa.

Klasifikasi merupakan suatu proses yang dapat digunakan untuk mengaitkan input dengan kategori atau label yang spesifik, seperti misalnya menilai apakah catatan akademis seorang siswa dapat dikategorikan sebagai lulus atau tidak lulus [5]. Upaya memprediksi kelulusan mahasiswa bertujuan membantu mahasiswa dalam mengetahui status kelulusannya. Dengan mengetahui prediksi durasi masa studi, mahasiswa dapat meraih manfaat dengan lebih memperhatikan perjalanan studi pribadinya. Secara umum, hal ini juga bermanfaat bagi program studi untuk melakukan tindakan preventif terkait durasi studi masing-masing mahasiswa.

Berdasarkan penjelasan sebelumnya, Penelitian ini bertujuan untuk membuat sistem yang mampu mengklasifikasikan kelulusan mahasiswa berdasarkan evaluasi kinerja mereka pada tahun pertama dan kedua. Teknik data mining diterapkan untuk mengidentifikasi pola-pola kelulusan dari mahasiswa yang telah lulus, yang kemudian dijadikan dasar untuk memprediksi kelulusan mahasiswa di masa mendatang. Kemajuan pesat dalam bidang data mining selalu terkait dengan perkembangan teknologi informasi yang memungkinkan pengolahan data dalam skala besar. Oleh karena itu, dengan signifikansinya dalam konteks data mining, terdapat berbagai algoritma yang digunakan untuk mengolah big data, salah satunya melalui penerapan algoritma klasifikasi. [6]. Klasifikasi merupakan suatu proses yang dapat digunakan untuk mengaitkan input dengan kategori atau label yang spesifik, seperti misalnya menilai apakah catatan akademis seorang siswa dapat dikategorikan sebagai lulus atau tidak lulus. Beberapa algoritma yang sesuai untuk melakukan klasifikasi data meliputi SVM, NBC, K-NN, Discriminant Analysis, Decision Tree, dan Neural Networks. [5].

Algoritma Naive Bayes adalah salah satu algoritma yang digunakan untuk teknik klasifikasi data. Naive Bayes adalah metode klasifikasi yang menggunakan pendekatan probabilistik dan statistik, diperkenalkan oleh Thomas Bayes, seorang ilmuwan asal Inggris. Algoritma ini memprediksi kemungkinan kejadian di masa depan berdasarkan data historis, dengan pendekatan yang dikenal sebagai teorema

Bayes. Istilah "naive" digunakan karena algoritma ini mengasumsikan bahwa semua atribut bersifat independen satu sama lain. Naive Bayes merupakan algoritma pembelajaran induktif yang sangat efisien dan efektif dalam konteks machine learning dan data mining, meskipun didasarkan pada asumsi bahwa atribut-atribut tidak memiliki ketergantungan satu sama lain, performa Naive Bayes tetap kompetitif dalam proses klasifikasi. Meskipun keadaan di mana asumsi ini terlanggar jarang terjadi dalam data sebenarnya, performa klasifikasi Naive Bayes tetap tinggi. [7]. Tetapi, Naive Bayes memiliki kelemahan dalam hal sensitivitas terhadap jumlah fitur yang berlebihan, yang dapat menyebabkan penurunan akurasi klasifikasi. [8].

Rendahnya tingkat kelulusan tepat waktu di Fakultas Ilmu Komputer Universitas Al Asyariah Mandar. Untuk mengatasi masalah tersebut, penulis mengusulkan solusi berupa pembangunan sistem prediksi kelulusan mahasiswa dengan memanfaatkan algoritma Naive Bayes. Melalui penerapan sistem ini, diharapkan pihak kampus dapat melakukan intervensi proaktif dan memberikan dukungan yang lebih tepat pada mahasiswa yang berpotensi mengalami kesulitan. Melalui pendekatan ini, penelitian ini berupaya memberikan kontribusi positif terhadap efisiensi akademik dan pengembangan mahasiswa di Prodi Teknik Informatika.

2. TINJAUAN PUSTAKA

2.1. Mahasiswa

Istilah mahasiswa mengacu pada seseorang yang secara resmi terdaftar di lembaga pendidikan tinggi, baik itu negeri maupun swasta, untuk mengejar pendidikan. [1]

2.2. Kelulusan Tepat Waktu

setiap awal tahun akademik, jumlah mahasiswa yang diterima terus meningkat, namun tidak semua dari mereka dapat mengerjakan proses belajar sesuai dengan batas yang ditentukan, sehingga mengakibatkan penambahan data mahasiswa. [9].

2.3. Data mining

Data mining adalah proses pencarian pola atau informasi penting dalam data yang dipilih menggunakan teknik atau metode khusus. Teknik, metode, atau algoritma dalam data mining memiliki beragam jenis yang berbeda-beda.

2.4. Naive bayes

Naive Bayes adalah sebuah metode klasifikasi yang menggunakan pendekatan probabilitas dan statistik, diciptakan oleh ilmuwan Inggris bernama Thomas Bayes. Metode ini mengestimasi probabilitas kejadian di masa depan berdasarkan pengalaman sebelumnya, dan pendekatan ini dikenal sebagai hukum Bayes. Kata "naive" digunakan karena asumsi bahwa kondisi antar atribut tidak saling bergantung. [7]. Naive Bayes adalah salah satu teknik pembelajaran induktif yang sangat efektif dan efisien dalam konteks machine learning dan eksplorasi data. Persamaan dari teorema Bayes adalah :

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

2.5. Atribut

Tabel 1. Atribut

Gender	Usia	Program studi	IPS Semester 1 sampai 5	IPK semester 5	Lama studi
Pria / wanita	Usia mahasiswa	T.I / S.I	Nilai IPS	IPK	Tepat waktu / tidak tepat waktu

2.6. Google Colaboratory

Google Colaboratory adalah sebuah platform komputasi yang dibuat oleh Google. Platform ini menggunakan Jupyter notebook dan termasuk dalam kategori bahasa pemrograman tingkat tinggi. Python, sebagai bahasa pemrograman utamanya, mudah diimplementasikan, memiliki fleksibilitas untuk dikembangkan, dan dilengkapi dengan berbagai pustaka seperti Numpy, OpenCV, Tensorflow, dan Keras.

3. METODE PENELITIAN

A. Alat Dan Bahan Penelitian

Pada penelitian ini alat yang digunakan

1) Perangkat keras

Lapto dengan spesifikasi laptop asus, Intel(R) Celeron(R) N4000, RAM 4,00 GB.

2) Perangkat lunak

Google Colaboratory

Google Colaboratory adalah platform komputasi yang dikembangkan oleh Google. Platform ini menggunakan Jupyter Notebook atau IPython (Python interaktif) dan termasuk dalam kategori bahasa pemrograman tingkat tinggi. Google Drive

Google Drive merupakan layanan penyimpanan awan dari Google yang menawarkan ruang penyimpanan gratis hingga 15GB.

Microsoft Excel

Microsoft Excel merupakan aplikasi dalam paket Microsoft Office yang digunakan untuk memproses angka (aritmatika).

3) Bahan penelitian

Berikut adalah penjelasan lebih detail mengenai bahan penelitian yang disebutkan:

Data historis akademik mahasiswa

Data akademik yang digunakan dalam penelitian ini mencakup informasi tentang Indeks Prestasi Semester (IPS) mahasiswa dari semester 1 sampai 5, dan IPK. Data ini diperoleh dari sistem informasi akademik universitas.

Data demografis

Selain data akademik, penelitian ini juga mempertimbangkan faktor demografis mahasiswa, termasuk jenis kelamin dan usia.

Literatur dan penelitian terkait

Untuk mendukung landasan teoritis penelitian, literatur ilmiah dan penelitian terkait menjadi bahan acuan. Akses ke

perpustakaan digital, jurnal ilmiah, dan literatur terkait diperlukan untuk memahami konteks penelitian serta mendukung argumen dan interpretasi hasil.

B. Waktu dan Tempat Penelitian

Penelitian ini dilakukan Pada Februari sampai Mei 2024, di Universitas Al Asyariah Mandar Fakultas Ilmu Komputer, Kecamatan Polewali, Kabupaten Polewali Mandar, Provinsi Sulawesi Barat.

C. Tahapan Penelitian

1) Investigasi

Mengidentifikasi masalah berarti peneliti memulai dengan merumuskan permasalahan yang akan diteliti. Langkah ini sangat krusial karena perumusan masalah akan menjadi panduan utama sepanjang penelitian. Tanpa rumusan masalah yang jelas, peneliti dapat kehilangan arah dalam proses penelitian [10].

Investigasi merupakan tahap awal dari penelitian ini. Pada tahap ini, dilakukan identifikasi mendalam terhadap area spesifik yang berkaitan dengan prediksi kelulusan mahasiswa tepat waktu. Fokus utama ditujukan pada klarifikasi tujuan dan pertanyaan penelitian sebagai dasar perjalanan penelitian ini.

2) Studi literatur

Studi literatur menjadi langkah berikutnya yang melibatkan tinjauan komprehensif terhadap literatur ilmiah yang relevan dengan faktor-faktor kelulusan mahasiswa. Dalam tahap ini, landasan teoritis dan konseptual dibangun dari review literatur untuk mendukung struktur penelitian secara menyeluruh.

3) Penentuan model pengukuran dan sampel

Penentuan model pengukuran dan sampel merupakan tahap ketiga dalam penelitian ini. Pada tahap ini, variabel yang diukur, seperti Indeks Prestasi Semester (IPS), IPK, gender, umur, dan program studi, ditetapkan. Selain itu, model atau metode pengukuran yang paling sesuai dipilih, dan sampel penelitian ditentukan untuk mencapai representasi yang akurat.

4) Pengembangan Model Prediksi

Pengembangan model prediksi menjadi tahap berikutnya yang mencakup pengembangan model prediksi kelulusan tepat waktu dengan merinci atribut-atribut yang telah ditentukan sebelumnya. Pada tahap ini, algoritma prediktif, terutama Naive Bayes, digunakan untuk uji coba, pelatihan, dan penyesuaian model guna mencapai hasil yang optimal.

5) Pengumpulan data

Pengumpulan data menjadi fokus pada tahap berikutnya dengan merancang prosedur pengumpulan data menggunakan metode data administratif, dan studi pustaka.

6) Pengelolaan dan analisis data

Pengelolaan dan analisis data merupakan tahap penting yang mencakup pengelolaan data termasuk pembersihan, penyusunan, dan pengkodean. Proses analisis data dilakukan menggunakan metode yang sesuai, khususnya algoritma Naive Bayes, untuk mendapatkan hasil yang relevan dengan tujuan penelitian.

7) Penyusunan laporan akhir

Tahap terakhir adalah penyusunan laporan akhir yang mencakup semua tahapan penelitian dari tujuan hingga temuan. Pada tahap ini, interpretasi hasil analisis data dan implikasi penelitian disajikan secara rinci, sementara rekomendasi berdasarkan temuan memberikan arah untuk penelitian masa depan.

D. Teknik Pengumpulan Data

Proses pengumpulan data dilakukan melalui metode dokumentasi dengan mengambil data akademik dan demografis mahasiswa di Universitas Al Asyariah Mandar [11].

1) Metode Dokumentasi

Atribut seperti "Program Studi", "Gender", "IPS Semester 1-5", "IPK", "Usia", dan "Lama Studi" diperoleh dari data administratif kampus. Informasi ini diambil dari catatan akademis, sistem informasi mahasiswa, dan sumber administratif lainnya yang dimiliki oleh institusi.

2) Studi Pustaka

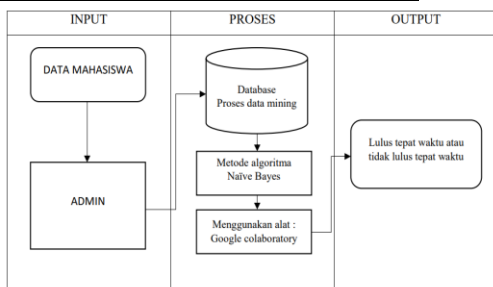
Informasi dan konsep terkait dengan faktor yang memengaruhi keberhasilan kelulusan mahasiswa metode-metode prediksi kelulusan, dan teori-teori terkait diperoleh melalui studi pustaka. Literatur ilmiah, jurnal-jurnal, buku, dan sumber-sumber akademis lainnya digunakan untuk membangun dasar teoritis dan konseptual penelitian.

E. Teknik analisis data

Penelitian ini mengadopsi metode kuantitatif dengan jenis penelitian eksperimental, di mana hipotesis diuji menggunakan pendekatan statistik. Penelitian ini melibatkan pemeriksaan beberapa variabel dengan menggunakan uji khusus dan mengaitkannya dengan permasalahan penelitian yang sedang diselidiki [12].

F. Kerangka Sistem

Nilai IPS & IPK	Skala Numerik
< 2,95	0
2,96 – 3,26	1
3,27 – 3,49	2
3,50 – 3,71	3
3,72 – 4,00	4



Gambar 1. Kerangka system

4. HASIL PENELITIAN

Berikut penulis tampilkan data yang digunakan dalam penelitian ini

Table 2. Tabel data penelitian

NIM	Nama	Program studi	Status mahasiswa
20180510070	Hariah	System informasi	Aktif

IPS	Jumlah SKS semester	IPK	Jumlah SKS total	Biaya kuliah
3,50	16	3,50	16	1

Data yang digunakan dalam penelitian ini adalah data mahasiswa dari angkatan 2018 dan 2019, jurusan teknik Informatika dan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Al Asyariah Mandar. Data keseluruhan yang didapat dari Fakultas Ilmu Komputer sebanyak 379 data.

Pembersihan data adalah proses membersihkan data dari kesalahan atau sinkonsistensi yang mungkin ada dalam dataset. Ini adalah langkah penting untuk memastikan kualitas data yang baik sebelum analisis dilakukan.

Setelah dilakukan pembersihan data diperoleh data sebanyak 58 data mahasiswa yang saip digunakan.

Tabel 3. Data hasil pembersihan

NPM	Nama	Gender	Usia	Program studi
201805100073	Muhammad Irsyad	L	22	System Informasi

IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPK	Status kelulusan
3,57	3,90	3,42	3,41	3,46	3,55	0

Transformasi data adalah proses mengubah bentuk atau karakteristik data dari format atau struktur awalnya menjadi format atau struktur lain yang lebih cocok untuk analisis atau pemodelan.

Tabel 4. Transformasi

Kategori gender	Skala Numerik
Laki-laki	1
Perempuan	0

Kategori	Skala Numerik
Lulus tepat waktu	1
Tidak lulus tepat waktu	0

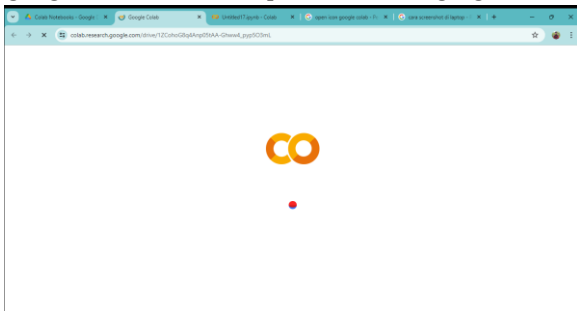
Tabel 5. Data transformasi

Gender	Usia	Program studi	IPS 1	IPS 2
0	20	0	4	3

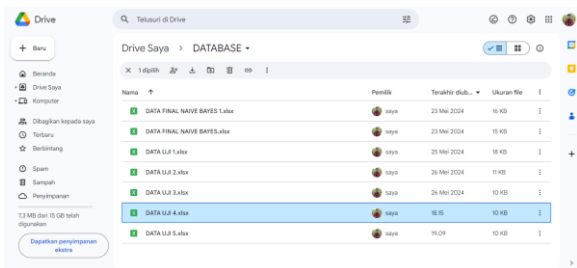
IPS 3	IPS 4	IPS 5	IPK	Status kelulusan
4	4	4	4	1

Pada langka ini, dilakukan pembuatan model data menggunakan metode prediksi dengan algoritma Naïve Bayes. Data yang telah dikumpulkan, dipilih, dan diubah akan dijalan dengan metode prediksi ini. Metode tersebut digunakan untuk memprediksi kemungkinan di waktu mendatang berdasarkan pengalaman sebelumnya sebagai perbandingan. Data yang akan diuji dibagi menjadi dua bagian, yakni data latihan dan data pengujian., kemudian dianalisis menggunakan Google Colab. Data training mahasiswa terdiri dari 40 record, sedangkan data testing terdiri dari 18 record.

Variabel yang dipakai sebagai label adalah status kelulusan, peneliti akan melakukan proses data mining menggunakan algoritma nive bayes menggunakan google colaboratori, sebelum melakukan pembuatan model, data yang digunakan harus di upload kedalam google drive.

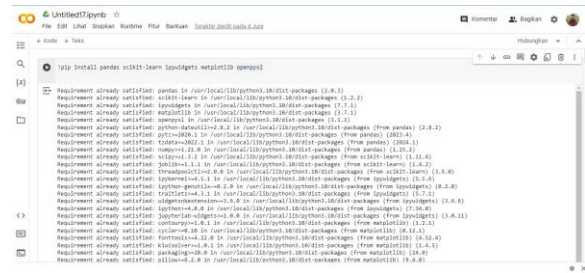


Gambar 2. Tampilan UI Google Colaboratory



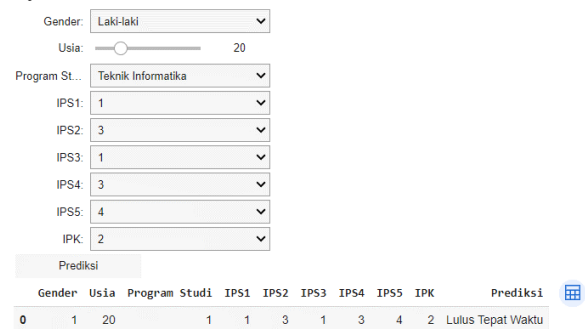
Gambar 3. Tampilan utama Google Drive

Google Drive merupakan layanan penyimpanan awan dari Google yang menawarkan ruang penyimpanan gratis hingga 15GB. Google drive digunakan sebagai database.



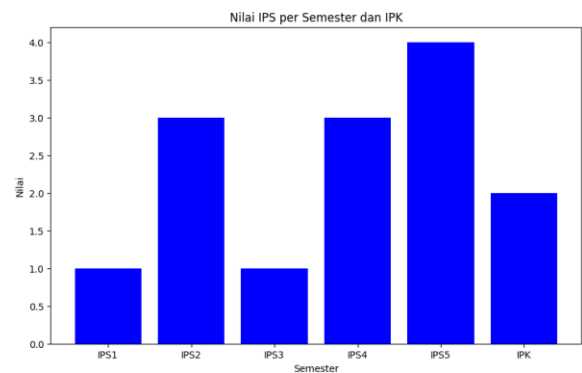
Gambar 4. Tampilan google colaboratory

Setelah melakukan inport data kedalam google drive, selanjutnya dilakukan pembuatan model prediksi di dalam google colaboratory. melakukan penginstalan beberapa paket Python menggunakan pip, yang merupakan package manager untuk Python.



Gambar 5. Tampilan widget input dan output

Semua widget input, tombol, dan output ditampilkan dalam satu kotak vertikal ('VBox'). Antarmuka ini memungkinkan interaksi yang mudah dan terstruktur, sehingga pengguna dapat memasukkan data, melihat hasil prediksi, dan mengeksplorasi visualisasi data dan laporan hasil klasifikasi.



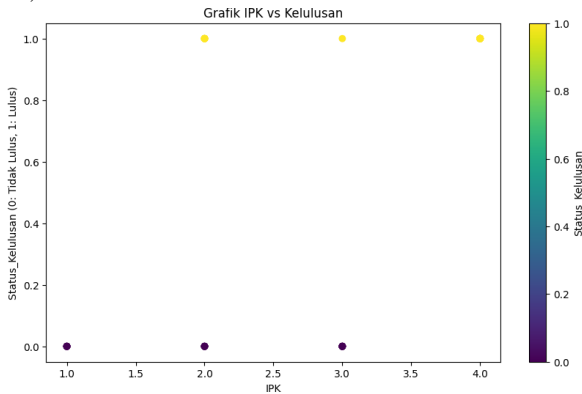
Gambar 6. Grafik nilai IPS dan IPK

Grafik ini ditampilkan di bawah hasil prediksi dan menunjukkan nilai IPS untuk setiap semester (IPS1 hingga IPS5) serta nilai IPK. Grafik ini memberikan visualisasi tentang bagaimana nilai-nilai tersebut tersebar dan apakah ada pola tertentu yang terlihat.

Mounted at /content/drive					
	precision	recall	f1-score	support	
Tidak Lulus Tepat Waktu	1.00	0.75	0.86	12	
Lulus Tepat Waktu	0.67	1.00	0.80	6	
accuracy			0.83	18	
macro avg	0.83	0.88	0.83	18	
weighted avg	0.89	0.83	0.84	18	

Gambar 7. Laporan hasil klasifikasi

Hasil klasifikasi yang dihasilkan oleh model Naïve Bayes. Laporan ini berisi metrik evaluasi model seperti precision, recall, dan F1-score.



Gambar 8. Grafik scatter plot

grafik scatter plot yang menunjukkan hubungan antara IPK dan status kelulusan untuk seluruh data dalam dataset.

Algoritma Naïve Bayes

Dalam studi ini, metode yang diterapkan adalah Naïve Bayes. Untuk mempermudah penjelasan tentang metode Naïve Bayes penulis menyusun tabel data yang digunakan dalam penelitian, termasuk tabel data latihan dan tabel data tes.

Tabel data training berisi 40 record yang digunakan sebagai data training, sedangkan tabel data testing berisi 18 record yang digunakan sebagai data testing dalam penelitian ini. Berdasarkan tabel tersebut, prediksi akan dibuat untuk menentukan apakah mahasiswa akan lulus tepat waktu atau tidak menggunakan teknik peluang. Untuk mengestimasi nilai peluang atau prediksi mahasiswa, digunakan langkah-langkah berikut :

$$P(H|X) = \frac{P(H|X) \cdot P(H)}{P(X)}$$

Langka pertama menghitung likelihood untuk setiap fitur dengan kelas "Lulus Tepat Waktu" dan "Terlambat Lulus" Prior Probability

$$\text{Prior(Lulus Tepat Waktu)} = 25/40 = 0.625$$

$$\text{Prior(Terlambat Lulus)} = 15/40 = 0.375$$

Gender

$$\text{Likelihood (Lakilaki | Lulus Tepat Waktu)} = 8/25 = 0.32$$

$$\text{Likelihood (Perempuan | Lulus Tepat Waktu)} = 17/25 = 0.68$$

$$\text{Likelihood (Lakilaki | Terlambat Lulus)} = 8/15 = 0.533$$

$$\text{Likelihood (Perempuan | Terlambat Lulus)} = 7/15 = 0.467$$

Usia

$$\text{Likelihood (Usia 18 | Lulus Tepat Waktu)} = 2/25 = 0.08$$

$$\text{Likelihood (Usia 19 | Lulus Tepat Waktu)} = 7/25 = 0.28$$

$$\text{Likelihood (Usia 20 | Lulus Tepat Waktu)} = 11/25 = 0.44$$

$$\text{Likelihood (Usia 21 | Lulus Tepat Waktu)} = 4/25 = 0.16$$

$$\text{Likelihood (Usia 22 | Lulus Tepat Waktu)} = 0/25 = 0$$

$$\text{Likelihood (Usia 23 | Lulus Tepat Waktu)} = 1/25 = 0.04$$

$$\text{Likelihood (Usia 29 | Lulus Tepat Waktu)} = 0/25 = 0$$

$$\text{Likelihood (Usia 38 | Lulus Tepat Waktu)} = 0/25 = 0$$

$$\text{Likelihood (Usia 18 | Terlambat Lulus)} = 0/15 = 0$$

$$\text{Likelihood (Usia 19 | Terlambat Lulus)} = 5/15 = 0.333$$

$$\text{Likelihood (Usia 20 | Terlambat Lulus)} = 5/15 = 0.333$$

$$\text{Likelihood (Usia 21 | Terlambat Lulus)} = 3/15 = 0.2$$

$$\text{Likelihood (Usia 22 | Terlambat Lulus)} = 0/15 = 0$$

$$\text{Likelihood (Usia 23 | Terlambat Lulus)} = 0/15 = 0$$

$$\text{Likelihood (Usia 29 | Terlambat Lulus)} = 1/15 = 0.067$$

$$\text{Likelihood (Usia 38 | Terlambat Lulus)} = 1/15 = 0.067$$

Program Studi

$$\text{Likelihood (Teknik Informatika | Lulus Tepat Waktu)} = 8/25 = 0.32$$

$$\text{Likelihood (Sistem Informasi | Lulus Tepat Waktu)} = 17/25 = 0.68$$

$$\text{Likelihood (Teknik Informatika | Terlambat Lulus)} = 4/15 = 0.267$$

$$\text{Likelihood (Sistem Informasi | Terlambat Lulus)} = 11/15 = 0.733$$

IPS1

$$\text{Likelihood (IPS1 = 0 | Lulus Tepat Waktu)} = 2/25 = 0,08$$

$$\text{Likelihood (IPS1 = 1 | Lulus Tepat Waktu)} = 6/25 = 0,24$$

$$\text{Likelihood (IPS1 = 2 | Lulus Tepat Waktu)} = 0/25 = 0$$

$$\text{Likelihood (IPS1 = 3 | Lulus Tepat Waktu)} = 8/25 = 0,32$$

$$\text{Likelihood (IPS1 = 4 | Lulus Tepat Waktu)} = 9/25 = 0,36$$

$$\text{Likelihood (IPS1 = 0 | Terlambat Lulus)} = 5/15 = 0.4$$

$$\text{Likelihood (IPS1 = 1 | Terlambat Lulus)} = 3/15 = 0.2$$

Likelihood (IPS1 = 2 | Terlambat Lulus) = $1/15 = 0.066$

Likelihood (IPS1 = 3 | Terlambat Lulus) = $3/15 = 0.2$

Likelihood (IPS1 = 4 | Terlambat Lulus) = $3/15 = 0.2$

IPS2

Likelihood (IPS2 = 0 | Lulus Tepat Waktu) = $0/25 = 0$

Likelihood (IPS2 = 1 | Lulus Tepat Waktu) = $1/25 = 0.04$

Likelihood (IPS2 = 2 | Lulus Tepat Waktu) = $1/25 = 0.04$

Likelihood (IPS2 = 3 | Lulus Tepat Waktu) = $13/25 = 0.52$

Likelihood (IPS2 = 4 | Lulus Tepat Waktu) = $10/25 = 0.4$

Likelihood (IPS2 = 0 | Terlambat Lulus) = $1/15 = 0.066$

Likelihood (IPS2 = 1 | Terlambat Lulus) = $5/15 = 0.333$

Likelihood (IPS2 = 2 | Terlambat Lulus) = $2/15 = 0.133$

Likelihood (IPS2 = 3 | Terlambat Lulus) = $5/15 = 0.333$

Likelihood (IPS2 = 4 | Terlambat Lulus) = $2/15 = 0.133$

IPS3

Likelihood (IPS3 = 0 | Lulus Tepat Waktu) = $0/25 = 0$

Likelihood (IPS3 = 1 | Lulus Tepat Waktu) = $3/25 = 0.12$

Likelihood (IPS3 = 2 | Lulus Tepat Waktu) = $5/25 = 0.2$

Likelihood (IPS3 = 3 | Lulus Tepat Waktu) = $1/25 = 0.04$

Likelihood (IPS3 = 4 | Lulus Tepat Waktu) = $16/25 = 0.64$

Likelihood (IPS3 = 0 | Terlambat Lulus) = $2/15 = 0.133$

Likelihood (IPS3 = 1 | Terlambat Lulus) = $2/15 = 0.133$

Likelihood (IPS3 = 2 | Terlambat Lulus) = $3/15 = 0.2$

Likelihood (IPS3 = 3 | Terlambat Lulus) = $6/15 = 0.4$

Likelihood (IPS3 = 4 | Terlambat Lulus) = $2/15 = 0.133$

IPS4

Likelihood (IPS4 = 0 | Lulus Tepat Waktu) = $0/25 = 0$

Likelihood (IPS4 = 1 | Lulus Tepat Waktu) = $1/25 = 0.04$

Likelihood (IPS4 = 2 | Lulus Tepat Waktu) = $2/25 = 0.08$

Likelihood (IPS4 = 3 | Lulus Tepat Waktu) = $8/25 = 0.32$

Likelihood (IPS4 = 4 | Lulus Tepat Waktu) = $16/25 = 0.68$

Likelihood (IPS4 = 0 | Terlambat Lulus) = $2/15 = 0.133$

Likelihood (IPS4 = 1 | Terlambat Lulus) = $4/15 = 0.266$

Likelihood (IPS4 = 2 | Terlambat Lulus) = $5/15 = 0.333$

Likelihood (IPS4 = 3 | Terlambat Lulus) = $2/15 = 0.133$

Likelihood (IPS4 = 4 | Terlambat Lulus) = $2/15 = 0.133$

IPS5

Likelihood (IPS5 = 0 | Lulus Tepat Waktu) = $0/25 = 0$

Likelihood (IPS5 = 1 | Lulus Tepat Waktu) = $0/25 = 0$

Likelihood (IPS5 = 2 | Lulus Tepat Waktu) = $1/25 = 0.04$

Likelihood (IPS5 = 3 | Lulus Tepat Waktu) = $8/25 = 0.32$

Likelihood (IPS5 = 4 | Lulus Tepat Waktu) = $16/25 = 0.64$

Likelihood (IPS5 = 0 | Terlambat Lulus) = $1/15 = 0.066$

Likelihood (IPS5 = 1 | Terlambat Lulus) = $3/15 = 0.2$

Likelihood (IPS5 = 2 | Terlambat Lulus) = $7/15 = 0.466$

Likelihood (IPS5 = 3 | Terlambat Lulus) = $3/15 = 0.2$

Likelihood (IPS5 = 4 | Terlambat Lulus) = $1/15 = 0.066$

IPK

Likelihood (IPK = 0 | Lulus Tepat Waktu) = $0/25 = 0$

Likelihood (IPK = 1 | Lulus Tepat Waktu) = $0/25 = 0$

Likelihood (IPK = 2 | Lulus Tepat Waktu) = $8/25 = 0.32$

Likelihood (IPK = 3 | Lulus Tepat Waktu) = $1/25 = 0.04$

Likelihood (IPK = 4 | Lulus Tepat Waktu) = $16/25 = 0.64$

Likelihood (IPK = 0 | Terlambat Lulus) = $0/15 = 0$

Likelihood (IPK = 1 | Terlambat Lulus) = $5/15 = 0.333$

Likelihood (IPK = 2 | Terlambat Lulus) = $4/15 = 0.266$

Likelihood (IPK = 3 | Terlambat Lulus) = $6/15 = 0.4$

Likelihood (IPK = 4 | Terlambat Lulus) = $0/15 = 0$

Selanjutnya Menghitung Posterior Probability dengan menggunakan data testing.

Tabel 6. Sampel data testing

Gender	Usia	Program studi	IPS 1	IPS 2
1	20	0	2	4
IPS 3	IPS 4	IPS 5	IPK	
2	1	3	2	

Sampel :

Prior (lulus tepat waktu) * Likelihood (Laki-laki | Lulus tepat waktu) * Likelihood (Usia 20 | Lulus tepat waktu) * Likelihood (S.I | Lulus tepat waktu) * Likelihood (IPS1 = 2 | Lulus tepat waktu) * Likelihood (IPS2 = 4 | Lulus tepat waktu) * Likelihood (IPS3 = 2 | Lulus tepat waktu) * Likelihood (IPS4 = 1 | Lulus tepat waktu) * Likelihood (IPS5 = 3 | Lulus tepat waktu) * Likelihood (IPK = 2 | Lulus tepat waktu) = $0.625 * 0.32 * 0.44 * 0.6 * 0 * 0.4 * 0.2 * 0.04 * 0.32 * 0.32 = 0$

Prior (Terlambat lulus) * Likelihood (Laki-laki | Terlambat lulus) * Likelihood (Usia 20 | Terlambat lulus) * Likelihood (S.I | Terlambat lulus) * Likelihood (IPS1 = 3 | Terlambat lulus) * Likelihood (IPS2 = 5 | Terlambat lulus) * Likelihood (IPS3 = 3 | Terlambat lulus) * Likelihood (IPS4 = 1 | Terlambat lulus) * Likelihood (IPS5 = 3 | Terlambat lulus) * Likelihood (IPK = 3 | Terlambat lulus) =
 $= 0.375 * 0.533 * 0.333 * 0.733 * 0.2 * 0.133 * 0.2 * 0.266 * 0.2 * 0.23$
 $= 3.175$

Posterior(Lulus Tepat Waktu) = 0

Posterior(Terlambat Lulus) = 3.175

Karena Posterior (Terlambat Lulus) lebih tinggi daripada Posterior (Lulus Tepat Waktu), maka prediksi untuk mahasiswa tersebut adalah "Terlambat Lulus".

5. KESIMPULAN

Hasil penelitian menunjukkan bahwa metode Naive Bayes menunjukkan kemampuan yang handal dalam memproyeksikan kelulusan mahasiswa tepat waktu, dengan tingkat ketepatan keseluruhan sebesar 83%. Model ini menunjukkan kinerja yang sangat baik dalam memprediksi mahasiswa yang tidak lulus tepat waktu, dengan nilai precision 1.00, recall 0.75, dan F1-score 0.86 untuk kelas "Tidak Lulus Tepat Waktu". Untuk prediksi mahasiswa yang lulus tepat waktu, model memiliki nilai precision 0.67, recall 1.00, dan F1-score 0.80. Secara keseluruhan, rata-rata makro (macro avg) menunjukkan precision 0.83, recall 0.88, dan F1-score 0.83. Sementara itu, rata-rata tertimbang (weighted avg) menunjukkan precision 0.89, recall 0.83, dan F1-score 0.84. Variabel-variabel seperti IPS dan IPK memberikan kontribusi signifikan terhadap prediksi kelulusan tepat waktu, sementara variabel gender, umur, dan program studi juga berperan dalam model prediksi meskipun dengan tingkat pengaruh yang berbeda-beda. Hasil ini menunjukkan bahwa model Naive Bayes efektif dalam mengidentifikasi mahasiswa yang berpotensi tidak lulus tepat waktu, dan memberikan wawasan yang berharga bagi pihak fakultas untuk mengambil tindakan preventif yang tepat.

6. DAFTAR PUSTAKA

- [1] J. Kurniawati and S. Baroroh, "Literasi Media Digital Mahasiswa Universitas Muhammadiyah Bengkulu," *J. Komun.*, vol. 8, no. 2, pp. 51–66, 2016.
- [2] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, p. 35, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [3] A. Putri *et al.*, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
- [4] K. H. L. Malelak, I. M. D. Ardiada, and G. Feoh, "Implementasi Klasifikasi Naive Bayes Dalam Memprediksi Lama Studi Mahasiswa (Studi Kasus : Universitas Dhyana Pura)," *SINTECH (Science Inf. Technol. J.*, vol. 4, no. 2, pp. 202–209, 2021, doi: 10.31598/sintechjournal.v4i2.964.
- [5] Y. P. Chiu, "Social recommendations for facebook brand pages," *J. Theor. Appl. Electron. Commer. Res.*, vol. 16, no. 1, pp. 71–84, 2021, doi: 10.4067/S0718-18762021000100106.
- [6] I. Parlina *et al.*, "Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012031.
- [7] Syarli and A. A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilm. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2016, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [8] D. A. Mutia, "Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *Jurnal Ilmu Pengetah. Dan Teknol. Komput.*, vol. 2, no. 2, pp. 39–45, 2017, [Online]. Available: <http://ejournal.nusamandiri.ac.id/index.php/jitk/article/view/381%0Ahttps://ejournal.nusamandiri.ac.id/index.php/jitk/article/download/381/337>
- [9] IRWAN, ADNAN SAUDDIN, and NUR IDA, "Penerapan Pohon Keputusan Dalam Memprediksi Masa Studi Mahasiswa Uin Alauddin Makassar," *J. INSTEK (Informatika Sains dan Teknol.*, vol. 7, no. 2, pp. 201–210, 2022, doi: 10.24252/instek.v7i2.31390.
- [10] N. Ridha, "PROSES PENELITIAN, MASALAH, VARIABEL DAN PARADIGMA PENELITIAN," *Comput. Graph. Forum*, vol. 39, no. 1, pp. 672–673, 2020, doi: 10.1111/cgf.13898.
- [11] S. Samasil, Y. Yuyun, and H. Hazriani, "Klasifikasi Mahasiswa Berpotensi Drop Out Menggunakan Algoritma Naive Bayes Dan Decision Tree," *J. Ilm. Ilmu Komput.*, vol. 8, no. 2, pp. 108–114, 2022, doi: 10.35329/jiik.v8i2.242.
- [12] A. Noviriandini and N. Nurajijah, "Analisis Kinerja Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Prestasi Siswa Sekolah Menengah Kejuruan," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 1, pp. 23–28, 2019, doi: 10.33480/jitk.v5i1.607.