



# Analisis Pemetaan Kemampuan Akademik Mahasiswa Program Studi Teknik Informatika Dengan Metode *Naive Bayes Classifier* Dan *K-Nearest Neighbor* (KNN) (Studi Kasus: Universitas Pamulang)

Utari Sekar Putri<sup>1</sup>, Taswanda Taryo<sup>2</sup>, Achmad Hindasyah<sup>3</sup>

<sup>1,2</sup>Program Studi Magister Teknik Informatika, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

<sup>3</sup>Badan Riset dan Inovasi Nasional (BRIN), Indonesia

Email: <sup>1</sup>utarisekarputri98@gmail.com, dosen02234@unpam.ac.id, achm021@brin.go.id

**Abstrak**– Pendidikan tinggi memiliki peran strategis dalam menghasilkan sumber daya manusia yang berkualitas, khususnya di bidang Teknik Informatika yang terus berkembang pesat. Penilaian performa akademik mahasiswa menjadi aspek penting untuk memastikan efektivitas pendidikan dan kualitas lulusan yang dihasilkan. Penelitian ini bertujuan untuk mengevaluasi proses pembelajaran mahasiswa Teknik Informatika dengan menerapkan teknik data mining, yaitu algoritma *Naive Bayes Classifier* (NBC) dan *K-Nearest Neighbors* (KNN). Metode penelitian yang digunakan melibatkan analisis data akademik mahasiswa Universitas Pamulang, yang mencakup tingkat keaktifan mahasiswa, pola kelulusan, serta faktor-faktor lain yang memengaruhi performa akademik. Proses analisis data dilakukan dengan membandingkan performa algoritma NBC dan KNN dalam memprediksi keberhasilan akademik mahasiswa berdasarkan parameter yang telah ditentukan. Selain itu, penelitian ini juga dilengkapi dengan survei yang bertujuan mengidentifikasi kendala yang memengaruhi tingkat kelulusan mahasiswa. Hasil penelitian menunjukkan bahwa metode KNN memiliki keunggulan dibandingkan NBC dengan akurasi masing-masing sebesar 91% untuk KNN dan 73% untuk NBC. Analisis data juga mengungkap adanya tren penurunan tingkat keaktifan mahasiswa dari 100% pada semester pertama menjadi 76,36% pada semester keempat, serta tingkat kelulusan tepat waktu yang sangat rendah, hanya 7,27% dari total populasi mahasiswa. Berdasarkan survei, kendala utama yang menyebabkan rendahnya tingkat kelulusan tepat waktu adalah masalah manajemen waktu, yang diakui oleh 71,43% responden, sementara beban tugas atau ujian tidak dianggap signifikan. Temuan ini menjadi dasar penting bagi pengelola program studi dalam merancang strategi pendidikan yang lebih responsif dan efektif untuk meningkatkan efisiensi studi serta kesiapan mahasiswa menghadapi tantangan dunia profesional.

**Kata Kunci:** Kelulusan Tepat waktu, *Naive Bayes*, *K-Nearest Neighbor*, Tingkat Keaktifan mahasiswa

**Abstract**– Higher education plays a strategic role in producing quality human resources, particularly in the rapidly developing field of Informatics Engineering. Evaluating students' academic performance is a crucial aspect to ensure the effectiveness of education and the quality of graduates. This research aims to evaluate the learning process of Informatics Engineering students by applying data mining techniques, namely the *Naive Bayes Classifier* (NBC) and *K-Nearest Neighbors* (KNN) algorithms. The research method involved analyzing academic data of Universitas Pamulang students, including their level of class activity, graduation patterns, and other factors affecting academic performance. The data analysis process compared the performance of NBC and KNN algorithms in predicting students' academic success based on predefined parameters. Additionally, the research was complemented by a survey aimed at identifying challenges affecting students' graduation rates. The research findings show that the KNN method outperformed NBC, with accuracy rates of 91% for KNN and 73% for NBC. Data analysis also revealed a declining trend in student participation from 100% in the first semester to 76.36% in the fourth semester, as well as a very low on-time graduation rate of only 7.27% of the total student population. According to the survey, the main challenge causing the low on-time graduation rate was time management issues, acknowledged by 71.43% of respondents, while task or exam workloads were not considered significant. These findings provide important insights for program managers in designing more responsive and effective educational strategies to improve study efficiency and better prepare students to face professional challenges.

**Keywords:** On-Time Graduation, *Naive Bayes*, *K-Nearest Neighbor*, Student Activity Level

## I. PENDAHULUAN

Pendidikan tinggi menjadi fokus utama dalam menghasilkan sumber daya manusia yang berkualitas dan mampu bersaing di era globalisasi saat ini. Program studi Teknik Informatika sebagai salah satu disiplin ilmu yang berkembang pesat, memiliki peran penting dalam mencetak para profesional yang handal di bidang teknologi. Seiring dengan kompleksitas perkembangan

teknologi, penilaian terhadap performa akademik mahasiswa menjadi krusial untuk mengevaluasi efektivitas pendidikan dan memastikan kualitas lulusan.[1]

Meskipun penilaian akademik mahasiswa Teknik Informatika telah dilakukan melalui berbagai metode seperti ujian tertulis, ujian praktik, dan penugasan, terdapat kesenjangan yang perlu diperhatikan untuk meningkatkan efektivitas proses penilaian. Salah satu



kesenjangan utama adalah belum adanya sistem pemantauan yang terintegrasi secara maksimal, sehingga perkembangan akademik mahasiswa tidak dapat terpantau dengan baik. Selain itu, partisipasi aktif mahasiswa dalam kelas, yang seharusnya menjadi indikator penting dalam menilai pemahaman mereka terhadap materi, sering kali tidak tercatat secara sistematis. Evaluasi akademik juga masih cenderung berfokus pada hasil ujian akhir tanpa memperhitungkan berbagai parameter lain seperti tugas harian dan keterampilan komunikasi. Hal ini menyebabkan potret kemampuan mahasiswa menjadi kurang komprehensif. Kurangnya pemanfaatan teknologi untuk mendukung pencatatan dan analisis data akademik turut memperburuk situasi, sehingga keputusan untuk perbaikan pembelajaran menjadi kurang berbasis data. Oleh karena itu, diperlukan langkah strategis untuk mengembangkan sistem penilaian yang lebih menyeluruh dan berbasis teknologi guna meningkatkan kualitas pendidikan serta memaksimalkan potensi mahasiswa

Di era digital dan kompetitif seperti sekarang, optimasi pengelolaan data akademik menjadi krusial untuk meningkatkan performa akademik mahasiswa. Salah satu cara yang dapat ditempuh adalah dengan menerapkan teknik data mining, khususnya menggunakan algoritma *Naive Bayes Classifier* dan *K-Nearest Neighbor (K-NN)*

Pemetaan kemampuan akademik sudah dilakukan dalam beberapa penelitian yang ada dengan berbagai penggunaan metode yang beragam. Seperti penelitian yang dilakukan oleh Arisa Dwi Cahyo[2] dengan meneliti Selama ini para mahasiswa belum dapat memprediksi nantinya apakah mereka akan lulus tepat ataupun terlambat, hal ini dapat mempengaruhi faktor akreditasi dari perguruan tinggi tersebut yang dapat menyebabkan beberapa kerugian baik dari mahasiswa ataupun dari pihak perguruan tinggi. Untuk memprediksi kelulusan mahasiswa tersebut menggunakan algoritma Naive Bayes. Pemilihan algoritma ini Naive Bayes dapat digunakan untuk melakukan prediksi probabilitas keanggotaan suatu class. Dengan adanya prediksi kelulusan diharapkan mampu memantu memprediksi tingkat kelulusan mahasiswa. Fitur – fitur yang digunakan dalam prediksi kelulusan mahasiswa yaitu Jenis Kelamin, Status Pernikahan, Status Pekerjaan, dan IPS1, IPS2, IPS3, IPS4. Prediksi kelulusan menggunakan data alumni yang sudah lulus. Berdasarkan perhitungan akurasi Naive Bayes menggunakan k-fold cross validation dengan hasil perhitungan rata-rata k=5 adalah sebesar 95%, dan perhitungan rata-rata k=10 sebesar 94% Berdasarkan hasil tersebut maka algoritma Naive Bayes bisa digunakan untuk prediksi kelulusan mahasiswa. Penelitian yang dilakukan cukup mirip dengan yang dilakukan oleh Arisa Dwi Cahyo dengan menggunakan metode Naive Bayes dan K-NN dan memfokuskan penelitian pada bidang prodi teknik informatika.

Naive Bayes merupakan algoritma yang digunakan untuk klasifikasi yang menggunakan teorema Bayes dan berasumsi bahwa nilai antar variabel saling bebas (independen) pada suatu nilai output. Dalam hal ini, diasumsikan bahwa kehadiran atau ketiadaan dari suatu

variabel tertentu tidak berhubungan dengan kehadiran atau ketiadaan dari variabel lainnya.[3]

Penelitian yang dilakukan oleh Wenti Dwi Yuniarti dan kawan-kawan [4] menunjukkan terobosan penting dalam evaluasi potensi keberhasilan studi mahasiswa melalui penerapan algoritma Naive Bayes Classifier. Dengan memanfaatkan empat variabel target, yaitu IPK Tahun Pertama dan Kedua, Status Mahasiswa Terkini, IPK Mata Kuliah (Makul) Non-Keprodian, serta IPK Makul Keprodian, penelitian ini memulai proses dengan tahap preprocessing yang menghasilkan 5.934 data bersih. Data tersebut kemudian dibagi menjadi 80% untuk training dan 20% untuk testing. Hasil analisis menunjukkan akurasi yang luar biasa, yaitu 99,41% untuk prediksi IPK Tahun Pertama dan Kedua, 96,96% untuk Status Mahasiswa Terkini, 95,87% untuk IPK Makul Keprodian, dan 97,89% untuk IPK Makul Non-Keprodian.

Keunggulan penelitian ini terletak pada kemampuan algoritma Naive Bayes dalam mengklasifikasikan data akademik dengan Correctly Classified Instances sebesar 97,53%, menunjukkan performa yang sangat tinggi dalam evaluasi akademik mahasiswa. Dengan hasil akurasi antara 95,8% hingga 99,41%, penelitian ini memberikan gambaran mendalam tentang efektivitas metode Naive Bayes untuk berbagai target evaluasi akademik. Temuan tersebut menegaskan bahwa penggunaan teknik data mining dapat mendukung perguruan tinggi dalam memprediksi potensi keberhasilan studi mahasiswa secara lebih akurat.

Lebih lanjut, penelitian ini menggarisbawahi pentingnya proses perekrutan mahasiswa yang lebih terarah serta perlunya memperhatikan faktor ekonomi orang tua sebagai salah satu penentu keberhasilan studi. Penelitian ini juga merintis penggunaan kombinasi algoritma Naive Bayes dan K-Nearest Neighbors (K-NN) yang diharapkan dapat mengevaluasi proses pembelajaran secara lebih efektif, mengurangi risiko over study, serta menekan potensi drop out mahasiswa. Kombinasi metode ini menawarkan solusi inovatif yang dapat diterapkan oleh institusi pendidikan tinggi dalam merancang strategi pembelajaran yang lebih responsif dan berbasis data.

Dengan demikian, Penilaian akademik yang efektif sangat diperlukan, tidak hanya untuk mengukur keberhasilan mahasiswa dalam proses pembelajaran, tetapi juga untuk memberikan wawasan yang mendalam kepada pengelola program studi dalam merumuskan kebijakan peningkatan kualitas pendidikan. Dalam konteks ini, teknik data mining, seperti Naive Bayes Classifier dan K-Nearest Neighbor (K-NN), telah terbukti mampu menghasilkan prediksi yang akurat dalam klasifikasi performa akademik mahasiswa di bidang Teknik Informatika. Dengan penerapan teknik ini, diharapkan pemetaan prestasi akademik mahasiswa dapat



lebih terstruktur, sehingga dosen dan pengelola program studi dapat melakukan evaluasi yang lebih tepat sasaran.

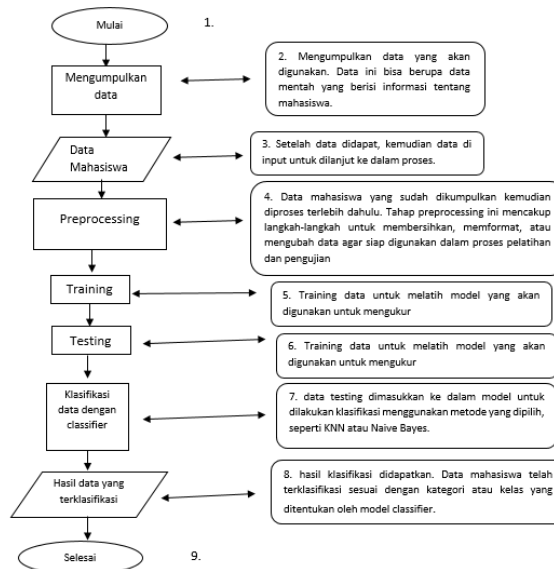
## II. METODE PENELITIAN

Penelitian ini mengadopsi metode dokumentasi sebagai salah satu teknik pengumpulan data. Melalui pendekatan ini, peneliti memanfaatkan berbagai dokumen yang tersedia di kampus, seperti laporan akademik, statistik jumlah mahasiswa, serta dokumen administrasi lainnya yang relevan dengan fokus penelitian ini. Data tersebut diperoleh dari sampel mata kuliah pada program studi Teknik Informatika S2.

Selain itu, untuk memperkaya data yang digunakan dalam penelitian ini, peneliti juga menyebarkan kuesioner kepada responden. Kuesioner tersebut dirancang untuk menggali informasi mendalam terkait berbagai faktor yang memengaruhi keaktifan dan keberhasilan akademik mahasiswa, termasuk manajemen waktu, beban tugas, serta tantangan akademik yang dihadapi. Hasil dari kuesioner ini kemudian diolah menggunakan algoritma Naive Bayes Classifier (NBC) dan K-Nearest Neighbors (KNN) untuk mengevaluasi efektivitas proses pembelajaran.

Sebagai bentuk penunjang analisis yang lebih komprehensif, hasil dari kuesioner dan proses pengolahan data ditampilkan dalam bentuk visualisasi diagram atau grafik. Visualisasi ini memberikan gambaran yang lebih jelas mengenai distribusi responden serta hubungan antara variabel-variabel yang dianalisis. Data yang diperoleh menunjukkan pola penurunan keaktifan mahasiswa dari semester pertama hingga semester keempat, serta tingkat kelulusan tepat waktu yang rendah. Analisis ini diharapkan dapat menjadi masukan bagi pengelola program studi dalam merancang strategi pendidikan yang lebih efektif dan berbasis data.

### A. PERANCANGAN PENELITIAN



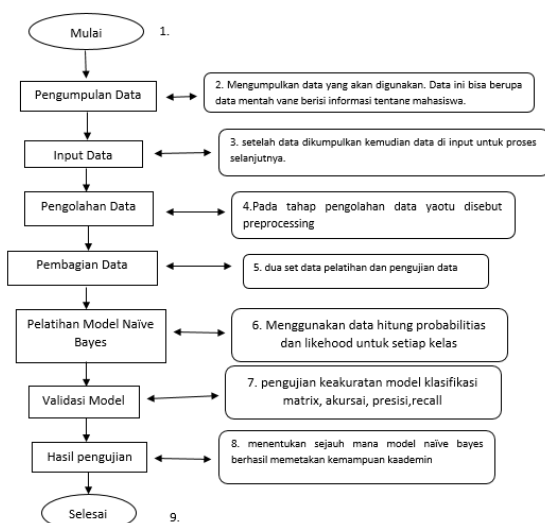
Gambar 1. Bagan Alur Rancangan Penelitian

Penjelasan Rancangan penelitian.

1. Mulai
2. Pengumpulan data yaitu data yang digunakan berupa dataset prodi teknik informatika S2 periode 2022 ganjil dan dataset dari kusioner, data yang digunakan sebanyak 55 data mahasiswa, data diperoleh dari hasil dokumentasi pihak akademik. Data diperlukan untuk proses input data untuk proses pengolahan data.
3. Setelah data didapat, kemudian data di input untuk dilanjut ke dalam proses, data yang telah di peroleh kemudian di convert ke dalam format CSV (Comma Separated Values) agar data dapat di proses dengan mudah menggunakan Phyton.
4. Data mahasiswa yang sudah dikumpulkan kemudian diproses terlebih dahulu. Tahap preprocessing ini mencakup langkah-langkah untuk membersihkan, memformat, atau mengubah data agar siap digunakan dalam proses pelatihan dan pengujian, Selain itu, dilakukan juga seleksi fitur untuk memastikan hanya atribut yang relevan digunakan dalam analisis. Preprocessing data ini sangat penting untuk meningkatkan kualitas data dan memastikan model dapat mempelajari pola dengan lebih akurat sehingga hasil analisis lebih optimal.
5. Training bagian data yang digunakan untuk melatih (training) model atau Data digunakan untuk melatih model agar dapat belajar pola dari data. Data ini diberikan kepada model dalam bentuk input (fitur) dan output (label) agar model dapat memahami dan memprediksi hasil berdasarkan pola yang ada. Selama proses pelatihan, algoritma akan mengoptimalkan parameter-parameter model untuk meminimalkan kesalahan dalam prediksi. Proses ini sangat penting untuk membangun model yang dapat secara akurat memproses data baru, sehingga model mampu melakukan generalisasi dan memberikan hasil yang relevan dalam pengujian atau implementasi.



6. Testing bagian data yang digunakan untuk menguji (testing) model yang sudah dilatih agar dapat mengukur performa model, Data digunakan untuk menguji performa model yang telah dilatih, sehingga hasil klasifikasinya dapat dievaluasi. Dengan menggunakan data testing, hasil klasifikasi model dapat dievaluasi melalui berbagai metrik, seperti akurasi, recall, precision, dan F1-score. Proses ini penting untuk memastikan bahwa model tidak hanya bekerja baik pada data latih tetapi juga mampu menangani data baru dengan tingkat kesalahan yang minimal.
7. Setelah model dilatih, data testing dimasukkan ke dalam model untuk dilakukan klasifikasi menggunakan metode yang dipilih, seperti KNN atau Naive Bayes. Pada tahap ini, model akan memproses data testing dan mengelompokkan setiap data ke dalam kelas yang sesuai berdasarkan pola yang telah dipelajari selama pelatihan. Algoritma klasifikasi akan membandingkan fitur-fitur dalam data testing dengan model yang telah terbentuk dan menghasilkan prediksi kategori atau kelas yang paling relevan. Hasil klasifikasi ini kemudian dapat digunakan untuk mengevaluasi kinerja model dan mengukur seberapa baik model tersebut dalam menggeneralisasi pola pada data yang belum pernah dilihat sebelumnya.
8. Setelah data diuji, hasil klasifikasi didapatkan. Data mahasiswa telah terklasifikasi sesuai dengan kategori atau kelas yang ditentukan oleh model classifier. Proses ini mengelompokkan data mahasiswa ke dalam kategori yang relevan, seperti misalnya status kelulusan tepat waktu atau tidak, berdasarkan pola yang telah dipelajari oleh model selama pelatihan. Hasil klasifikasi ini memberikan informasi yang berguna untuk analisis lebih lanjut dan evaluasi efektivitas model dalam mengidentifikasi pola dalam data, serta menjadi dasar untuk pengambilan keputusan atau rekomendasi yang lebih tepat.
9. Selesai.



**Gambar 2.** Bagan Alur Perancangan Sistem *Naive Bayes* Secara Umum

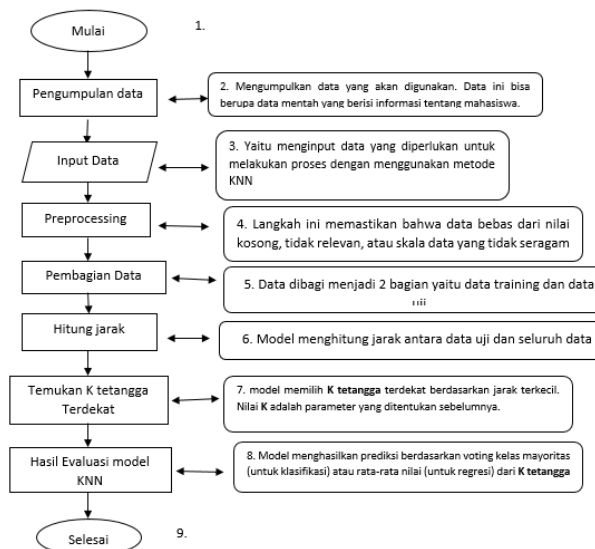
Penjelasan Perancangan sistem *naive bayes* secara umum:

1. Mulai dimana proses akan dimulai
2. Pengumpulan data yaitu data yang digunakan berupa dataset prodi teknik informatika S2 periode 2022 ganjil dan dataset dari kusioner, data yang digunakan sebanyak 55 data mahasiswa, data diperoleh dari hasil dokumentasi pihak akademik. Data diperlukan untuk proses pengolahan data
3. Setelah data didapat, kemudian data di input untuk dilanjut ke dalam proses, data yang telah di peroleh kemudian di *convert* ke dalam format *CSV (Comma Separated Values)* agar data dapat di proses dengan mudah menggunakan Phyton.
4. Data mahasiswa yang sudah dikumpulkan kemudian diproses terlebih dahulu. Tahap *preprocessing* ini mencakup langkah-langkah untuk membersihkan, memformat, atau mengubah data agar siap digunakan dalam proses pelatihan dan pengujian, Selain itu, dilakukan juga seleksi fitur untuk memastikan hanya atribut yang relevan digunakan dalam analisis. *Preprocessing* data ini sangat penting untuk meningkatkan kualitas data dan memastikan model dapat mempelajari pola dengan lebih akurat sehingga hasil analisis lebih optimal.
5. Pembagian Data Dataset dibagi menjadi dua bagian utama data pelatihan dan data pengujian. Data pelatihan digunakan untuk membangun model, di mana model akan belajar dari pola-pola yang ada dalam data tersebut. Selama proses pelatihan, model mengoptimalkan parameter untuk meminimalkan kesalahan prediksi. Sementara itu, data pengujian digunakan untuk mengevaluasi kinerja model setelah pelatihan, memastikan model dapat bekerja dengan baik pada data yang belum pernah dilihat sebelumnya. Pembagian ini penting untuk menghindari overfitting, sehingga model dapat menggeneralisasi dengan baik dan memberikan hasil yang akurat pada data baru.
6. Pelatihan Model *Naive Bayes* Metode ini bekerja dengan menghitung probabilitas prior dan likelihood untuk setiap kelas berdasarkan asumsi independensi antar fitur. Artinya, *Naive Bayes* menganggap bahwa setiap fitur dalam data bersifat independen satu sama lain, meskipun dalam kenyataannya hubungan antar fitur mungkin ada. Dalam proses ini, model menghitung distribusi probabilitas untuk setiap kelas, kemudian menggunakan aturan Bayes untuk memprediksi kelas data baru dengan mengkombinasikan prior probability dan likelihood yang telah dihitung. Pada kasus data kategori, model ini menggunakan *Categorical Naive Bayes* untuk menangani data dengan fitur yang bersifat kategorikal, memungkinkan model untuk memprediksi kelas berdasarkan distribusi probabilitas yang telah dilatih.
7. Validasi Model dilakukan setelah model dilatih untuk mengevaluasi kinerjanya menggunakan data pengujian. Proses ini melibatkan perhitungan berbagai metrik performa, seperti akurasi, presisi, recall, dan F1-score, yang digunakan untuk mengukur seberapa baik model dalam memprediksi kelas dengan benar. Akurasi mengukur proporsi prediksi yang benar dibandingkan dengan total data, sementara presisi



- menilai seberapa tepat prediksi positif yang dilakukan oleh model. Recall mengukur kemampuan model dalam mendeteksi semua data positif yang sebenarnya, dan F1-score memberikan keseimbangan antara presisi dan recall. Dengan menggunakan metrik-metrik ini, kita dapat menilai seberapa baik model dalam mengklasifikasikan data baru dan sejauh mana model dapat diterapkan pada kasus yang lebih luas.
8. Hasil Pengujian pada tahap ini, hasil dari validasi dianalisis untuk menentukan apakah model sudah cukup baik atau perlu penyempurnaan. Pengujian ini mencakup evaluasi matriks konfusi, yang menunjukkan distribusi prediksi yang benar dan salah dalam bentuk tabel perbandingan antara kelas yang diprediksi dan kelas yang sebenarnya. Selain itu, perhitungan akurasi digunakan untuk menilai proporsi prediksi yang benar secara keseluruhan. Analisis lebih mendalam terhadap kesalahan klasifikasi juga dilakukan untuk memahami jenis kesalahan yang terjadi, apakah model cenderung salah dalam mengklasifikasikan kelas tertentu, atau ada pola tertentu yang tidak terdeteksi. Dengan cara ini, kita dapat mengidentifikasi area yang perlu diperbaiki dan mengoptimalkan model untuk meningkatkan performa pada pengujian berikutnya.
  9. Selesai
- Setelah semua langkah di atas selesai, model Naïve Bayes dapat digunakan untuk memetakan kemampuan akademik mahasiswa atau untuk tujuan prediksi lainnya sesuai dengan data yang diberikan

3. Setelah data didapat, kemudian data di input untuk dilanjut ke dalam proses, data yang telah di peroleh kemudian di *convert* ke dalam format *CSV (Comma Separated Values)* agar data dapat di proses dengan mudah menggunakan *Phyton*
4. Data mahasiswa yang sudah dikumpulkan kemudian diproses terlebih dahulu. Tahap *preprocessing* ini mencakup langkah-langkah untuk membersihkan, memformat, atau mengubah data agar siap digunakan dalam proses pelatihan dan pengujian, Selain itu, dilakukan juga seleksi fitur untuk memastikan hanya atribut yang relevan digunakan dalam analisis. *Preprocessing* data ini sangat penting untuk meningkatkan kualitas data dan memastikan model dapat mempelajari pola dengan lebih akurat sehingga hasil analisis lebih optimal.
5. Pembagian Data Dataset dibagi menjadi dua bagian utama data pelatihan dan data pengujian. Data pelatihan digunakan untuk membangun model, di mana model akan belajar dari pola-pola yang ada dalam data tersebut. Selama proses pelatihan, model mengoptimalkan parameter untuk meminimalkan kesalahan prediksi. Sementara itu, data pengujian digunakan untuk mengevaluasi kinerja model setelah pelatihan, memastikan model dapat bekerja dengan baik pada data yang belum pernah dilihat sebelumnya. Pembagian ini penting untuk menghindari *overfitting*, sehingga model dapat menggeneralisasi dengan baik dan memberikan hasil yang akurat pada data baru.
6. Hitung Jarak, pada tahap ini, sistem menghitung jarak antara data uji dengan seluruh data latih untuk menentukan kemiripan atau kedekatan. Proses ini umumnya digunakan dalam algoritma seperti *K-Nearest Neighbors (KNN)*, di mana jarak dihitung menggunakan metrik tertentu, seperti *Euclidean* atau *Manhattan*, untuk mengukur seberapa mirip data uji dengan data latih. Jarak yang lebih pendek menunjukkan bahwa data uji lebih mirip dengan data latih tersebut, sementara jarak yang lebih panjang menunjukkan perbedaan yang lebih besar. Dengan menghitung jarak ini, model dapat menentukan kelas atau kategori yang paling sering muncul di antara data latih yang terdekat, yang kemudian digunakan untuk memprediksi kelas dari data uji.
7. Tetangga terdekat ini menjadi dasar untuk menentukan hasil prediksi atau klasifikasi. Temukan *K Tetangga Terdekat*, dari hasil perhitungan jarak, algoritma memilih *K tetangga terdekat* (nilai *K* ditentukan sebelumnya). *K tetangga* ini adalah data latih yang memiliki jarak paling dekat dengan data uji, berdasarkan metrik yang telah dihitung sebelumnya. Nilai *K* adalah parameter yang penting, karena menentukan jumlah tetangga yang akan dipertimbangkan dalam proses prediksi. Semakin besar nilai *K*, semakin banyak data latih yang diikutsertakan dalam menentukan hasil prediksi, yang dapat membantu mengurangi *noise* atau fluktuasi dalam klasifikasi. Setelah *K tetangga terdekat* ditemukan, algoritma menghitung mayoritas kelas dari



**Gambar 3.** Bagan Alur Algoritma KNN

Penjelasan dari Gambar 3.3 diatas sebagai berikut:

1. Mulai  
Proses dimulai dengan menyiapkan data dan algoritma KNN untuk melakukan analisis atau klasifikasi.
2. Pengumpulan data yaitu data yang digunakan berupa dataset prodi teknik informatika S2 periode 2022 ganjil dan dataset dari kusioner, data yang digunakan sebanyak 55 data mahasiswa, data diperoleh dari hasil dokumentasi pihak akademik. Data diperlukan untuk proses input data untuk proses pengolahan data



tetangga tersebut, dan kelas yang paling sering muncul akan menjadi prediksi atau klasifikasi untuk data uji.

8. Hasil Evaluasi Model KNN Berdasarkan Tetangga Terdekat, untuk klasifikasi, hasil prediksi ditentukan dengan voting kelas mayoritas dari tetangga terdekat. Setiap tetangga memberikan suara untuk kelas yang sesuai, dan kelas yang memperoleh suara terbanyak akan dipilih sebagai hasil prediksi. Setelah model menghasilkan prediksi, evaluasi dilakukan menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Akurasi mengukur seberapa banyak prediksi yang benar dibandingkan dengan total data, sementara precision menunjukkan ketepatan prediksi positif. Recall mengukur kemampuan model dalam mendeteksi semua data positif yang sebenarnya, dan F1-score memberikan keseimbangan antara precision dan recall. Metrik-metrik ini digunakan untuk menilai sejauh mana model KNN dapat melakukan klasifikasi yang akurat dan efektif pada data uji.

9. Selesai  
Proses berakhir dengan hasil prediksi atau klasifikasi yang dihasilkan oleh algoritma KNN

### B. PEMODELAN

Algoritma Naive Bayes Classifier yaitu algoritma klasifikasi berbasis teori probabilitas Bayes yang mengasumsikan bahwa fitur-fitur yang digunakan untuk klasifikasi bersifat independen (tidak saling bergantung). Dengan asumsi ini, Naive Bayes menghitung probabilitas setiap kelas berdasarkan data input dan memilih kelas dengan probabilitas tertinggi.

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

dengan :

X= Data dengan class yang belum diketahui

H= Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

1. Algoritma K-Nearest Neighbor yaitu algoritma klasifikasi yang berbasis pada jarak antara titik data. Dalam KNN, sebuah data akan diklasifikasikan ke dalam kelas mayoritas dari K tetangga terdekat yang paling mirip dengannya berdasarkan suatu metrik jarak (misalnya, Euclidean distance).

$$(xi, xj) = \sqrt{\sum_{r=1}^n (a_r(X_i) - a_r(X_j))^2}$$

Evaluasi Model

2. Akurasi: Akurasi mengukur seberapa banyak prediksi yang benar dibandingkan dengan total prediksi yang dilakukan. Ini adalah metrik yang paling umum digunakan untuk mengevaluasi kinerja model klasifikasi.
3. Presisi: Presisi mengukur seberapa banyak dari prediksi positif yang benar-benar relevan (positif

sejati) dibandingkan dengan semua prediksi yang dikategorikan positif oleh model.

4. *Recall*: *Recall* mengukur seberapa banyak data positif yang benar-benar terdeteksi (terprediksi) oleh model dibandingkan dengan total data positif yang sebenarnya ada.
5. F1-Score adalah rata-rata harmonis antara presisi dan recall. Metrik ini berguna ketika ada ketidakseimbangan antara presisi dan recall, karena memberikan nilai yang seimbang antara keduanya.
6. Analisis Hasil

Dari Algoritma Naive Bayes dan K-Nearest Neighbor (K-NN) dapat digunakan untuk memprediksi probabilitas kelulusan mahasiswa. Hasil evaluasi dapat menunjukkan akurasi tinggi dalam memprediksi apakah mahasiswa akan lulus tepat waktu atau mengalami keterlambatan dalam studi. Prediksi semacam ini membantu dosen dan pengelola program studi dalam mengambil langkah-langkah pencegahan dan pendukung agar mahasiswa dapat menyelesaikan studi dengan tepat waktu.

### III. HASIL DAN PEMBAHASAN

Berdasarkan implementasi klasifikasi dan prediksi yang telah dilakukan sebelumnya, proses pengujian dimulai dengan membagi dataset menjadi data latih (80%) dan data uji (20%) untuk memastikan model dapat dievaluasi secara objektif. Model KNN diimplementasikan dengan mempertimbangkan sejumlah tetangga terdekat yang optimal, sedangkan Naive Bayes memanfaatkan prinsip probabilitas untuk klasifikasi. Hasil pengujian menunjukkan performa model diukur berdasarkan akurasi, di mana prediksi dari kedua model dibandingkan dengan data aktual.

Hasil evaluasi menunjukkan bahwa metode KNN mencapai akurasi sebesar 91%, sedangkan Naive Bayes mencatat akurasi sebesar 73%. Sebagai data pembandingan, penelitian yang dilakukan oleh Wenti Dwi Yuniarti et al [4] . menunjukkan bahwa metode Naive Bayes mampu mencapai akurasi hingga 99,41% untuk prediksi IPK mahasiswa pada target tertentu. Perbedaan ini dapat terjadi karena variasi dalam jenis dataset, parameter input, serta karakteristik variabel target yang dianalisis.

SEMESTER 1 - Aktif: 100.00%   Tidak Aktif: 0.00%			
SEMESTER2 - Aktif: 83.64%   Tidak Aktif: 16.36%			
SEMESTER3 - Aktif: 78.18%   Tidak Aktif: 21.82%			
SEMESTER4 - Aktif: 76.36%   Tidak Aktif: 23.64%			
Persentase mahasiswa yang lulus tepat waktu: 7.27%			
Mahasiswa yang lulus tepat waktu:			
NIM	NAMA	Lulus	Tepat Waktu
1 22181200002	EKO KRISTIANO	Ya	Ya
23 221812000041	YOSSY VEIEBRIAN FITRI PRASHONO	Ya	Ya
41 221812000017	RINI HIDASWARI PURBA	Ya	Ya
45 221812000030	YUNUS PRAMONO	Ya	Ya
Akurasi model: 0.73			

Gambar 4. Hasil Pengujian Menggunakan Naive Bayes

Hasil pengujian pada Gambar 4. menggunakan metode *Naive Bayes* menunjukkan bahwa model mencapai akurasi sebesar 73%, yang berarti model berhasil memprediksi 73% data dengan benar berdasarkan data latih dan data uji. Analisis lebih lanjut terhadap dataset memperlihatkan bahwa persentase mahasiswa aktif pada Semester 1 mencapai 100%, menandakan seluruh mahasiswa terlibat aktif pada semester awal. Namun, tingkat keaktifan mulai menurun pada semester-semester berikutnya, yaitu 83.64% pada Semester 2, 78.18% pada Semester 3, dan 76.36% pada Semester 4,



dengan persentase mahasiswa tidak aktif meningkat secara bertahap. Selain itu, hanya 7.27% mahasiswa yang berhasil lulus tepat waktu, yaitu mahasiswa yang menyelesaikan studi sesuai dengan jadwal yang telah ditentukan. Nama-nama mahasiswa yang termasuk dalam kategori ini adalah Eko Kristianto, Yossy VeieBrian Fitri Prasmono, Rini Widawari Purba, dan Yunus Pramono, yang diidentifikasi berdasarkan atribut "Lulus Tepat Waktu" dalam dataset. Hasil ini menunjukkan bahwa meskipun akurasi model Naive Bayes cukup baik, tingkat kelulusan tepat waktu relatif rendah dibandingkan jumlah total mahasiswa.

Akurasi model KNN: 0.91

SEMESTER 1 - Aktif: 100.00% | Tidak Aktif: 0.00%

SEMESTER2 - Aktif: 83.64% | Tidak Aktif: 16.36%

SEMESTER3 - Aktif: 78.18% | Tidak Aktif: 21.82%

SEMESTER4 - Aktif: 76.36% | Tidak Aktif: 23.64%

Persentase mahasiswa yang lulus tepat waktu: 7.27%

Data mahasiswa yang lulus tepat waktu:

NIM	NAMA	Lulus Tepat Waktu
1 221012000002	EKO KRISTIANTO	Ya
23 221012000041	YOSSY VEIEBRIAN FITRI PRASMONO	Ya
41 221012000017	RINI WIDASWARI PURBA	Ya
45 221012000030	YUNUS PRAMONO	Ya

**Gambar 5.** Hasil pengujian menggunakan K-NN

Hasil pengujian pada gambar 4.30 menggunakan metode *K-Nearest Neighbors (KNN)* menunjukkan akurasi model sebesar 91%, yang berarti model memiliki tingkat keberhasilan tinggi dalam memprediksi data dengan benar. Analisis keaktifan mahasiswa memperlihatkan bahwa seluruh mahasiswa aktif pada Semester 1 (100%), namun persentase keaktifan mulai menurun pada semester berikutnya, yaitu 83.64% pada Semester 2, 78.18% pada Semester 3, dan 76.36% pada Semester 4. Sementara itu, persentase mahasiswa yang lulus tepat waktu hanya mencapai 7.27%, menunjukkan bahwa mayoritas mahasiswa tidak menyelesaikan studi sesuai jadwal yang ditentukan. Mahasiswa yang berhasil lulus tepat waktu teridentifikasi sebanyak empat orang, yaitu Eko Kristianto, Yossy VeieBrian Fitri Prasmono, Rini Widawari Purba, dan Yunus Pramono, yang tercatat dalam data dengan atribut "Lulus Tepat Waktu".

**Tabel 1.** Perbandingan metode Naive Bayes dan KNN dataset 1

Metode Naive Bayes	Accuracy	73%
	Recall	66%
	Precision	73%
	F1-Score	69%
Metode KNN	Accuracy	91%
	Recall	91%
	Precision	83%
	F1-Score	87%

Berdasarkan hasil evaluasi, metode *K-Nearest Neighbors (KNN)* menunjukkan kinerja yang lebih baik dibandingkan dengan metode *Naive Bayes*. *KNN* mencapai tingkat akurasi 91% dengan nilai recall dan F1-

score yang tinggi, masing-masing sebesar 91% dan 87%, yang menunjukkan kemampuan model untuk mendeteksi kelas dengan baik sekaligus menjaga keseimbangan antara presisi dan recall. Sementara itu, metode *Naive Bayes* memiliki akurasi sebesar 73% dengan nilai *F1-score* 69%, yang relatif lebih rendah dibandingkan *KNN*, terutama pada aspek recall sebesar 66%. Hal ini menunjukkan bahwa *KNN* lebih unggul dalam menangani dataset yang digunakan, baik dari segi prediksi maupun kemampuan generalisasi.

**Tabel 2.** Perbandingan Metode Naive Bayes dan KNN dataset 2

Metode Naive Bayes	Accuracy	72.73%
	Recall	75.00%
	Precision	85.71%
	F1-Score	80.00%
Metode KNN	Accuracy	54.55%
	Recall	37.50%
	Precision	100%
	F1-Score	54.55%

Berdasarkan hasil evaluasi performa model, Metode *Naive Bayes* menunjukkan kinerja yang lebih baik dibandingkan dengan Metode *KNN*. Dengan *Accuracy* sebesar 72.73%, *Naive Bayes* berhasil mengklasifikasikan data lebih tepat, disertai dengan *Recall* 75.00%, yang menunjukkan kemampuannya dalam mendeteksi sebagian besar kasus positif. *Precision-nya* yang mencapai 85.71% juga mengindikasikan bahwa model ini memiliki tingkat ketepatan yang tinggi dalam klasifikasi positif. *F1-Score* 80.00% mencerminkan keseimbangan yang baik antara *Precision* dan *Recall*. Di sisi lain, *KNN* memiliki *Accuracy* yang lebih rendah, yaitu 54.55%, dan *Recall* yang jauh lebih rendah, 37.50%, yang menunjukkan bahwa meskipun *KNN* memiliki *Precision* yang sempurna (100%), model ini kurang efektif dalam mendeteksi semua kasus positif. *F1-Score* sebesar 54.55% mengindikasikan bahwa meskipun *KNN* tepat dalam beberapa prediksi, keseimbangan antara *Precision* dan *Recall* masih perlu ditingkatkan. Secara keseluruhan, *Naive Bayes* lebih unggul dalam memberikan klasifikasi yang lebih tepat dan seimbang, sementara *KNN* memerlukan penyesuaian lebih lanjut untuk meningkatkan performanya dalam mengidentifikasi kategori positif.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode *K-Nearest Neighbors (KNN)* dengan akurasi 91% menunjukkan performa yang lebih baik dalam memprediksi data dibandingkan metode *Naive Bayes* yang memiliki akurasi 73%. Meskipun



demikian, Naive Bayes terbukti lebih unggul dalam mengevaluasi proses pembelajaran dengan keseimbangan yang baik antara recall, precision, dan F1-score, yang dapat memberikan analisis lebih komprehensif terkait keberhasilan proses akademik. Penurunan partisipasi mahasiswa dari Semester 1 hingga Semester 4 serta rendahnya tingkat kelulusan tepat waktu (hanya 7,27%) menjadi temuan signifikan yang menuntut perhatian serius dari institusi pendidikan. Hal ini menggarisbawahi perlunya strategi pendampingan akademik yang lebih efektif dan sistem pemantauan progres studi yang terstruktur. Penelitian ini memberikan kontribusi penting dalam pengembangan metode analisis akademik berbasis data untuk mendukung institusi dalam mengambil kebijakan yang lebih adaptif dan responsif, guna meningkatkan efisiensi studi dan memastikan mahasiswa lulus tepat waktu. Namun, tantangan berupa penurunan keaktifan dan rendahnya efisiensi studi perlu ditindaklanjuti dengan langkah-langkah yang lebih strategis dan berkelanjutan..

## REFERENSI

- [1] A. Febriyanto, S. Achmadi, and A. P. Sasmito, "Penerapan Metode K-Means Untuk Clustering Pengunjung Perpustakaan Itn Malang," *J. Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 61–70, 2021.
- [2] A. D. Cahyo, "Metode Naive Bayes Untuk Klasifikasi Masa Studi Sarjana," *J. Teknol. Pint.*, vol. 3, no. 4, 2023, [Online]. Available: <http://teknologipintar.org/index.php/teknologipintar/article/view/385%0Ahttp://teknologipintar.org/index.php/teknologipintar/article/download/385/370>
- [3] C. Journal, I. W. Saputro, B. W. Sari, P. Studi, I. Komputer, and J. Informatika, "Uji Performa Algoritma Naive Bayes untuk Prediksi Masa Studi Mahasiswa," vol. 6, no. 1, pp. 1–11, 2019.
- [4] W. D. Yuniarti, A. N. Faiz, and B. Setiawan, "Identifikasi Potensi Keberhasilan Studi Menggunakan Naive Bayes Classifier," *Walisongo J. Inf. Technol.*, vol. 2, no. 1, p. 1, 2020, doi: 10.21580/wjit.2020.2.1.5204.
- [5] B. D. Meilani, S. Wahyudiana, A. Y. P. Putri, and A. Pakarbudi, "Klasifikasi Identifikasi Faktor Penyebab Ketidaktepatan Masa Lulus Mahasiswa dengan Metode Naive Bayes Classifier," *Semin. Nas. Sains dan Teknol. Terap.*, pp. 297–302, 2019, [Online]. Available: <https://ejournal.itats.ac.id/sntekpan/article/view/586>
- [6] R. Hasudungan and W. J. Pranoto, "Implementasi Teorema Naive Bayes Pada Prediksi Prestasi Mahasiswa," vol. 5, no. 1, pp. 10–16, 2021.
- [7] D. Marutho, "PERBANDINGAN METODE NAIVE BAYES , KNN , DECISION TREE PADA LAPORAN WATER LEVEL JAKARTA," pp. 90–97, 2019.
- [8] A. Rahman, "Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes," vol. 9, pp. 22–31, 2023.
- [9] D. Ayu *et al.*, "AN EDUCATIONAL DATA MINING FOR STUDENT ACADEMIC PREDICTION USING K-MEANS CLUSTERING AND NAIVE BAYES," pp. 155–160.
- [10] A. F. Watratan, A. P. B, D. Moeis, S. Informasi, and S. P. Makassar, "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY ( JACOST ) Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," vol. 1, no. 1, pp. 7–14, 2020.
- [11] S. Iskandar, N. R. Refisis, and B. A. Ginting, "Metode Naive Bayes Classifier Dalam Penentuan Penerima Beasiswa Bidikmisi Di Universitas Negeri Medan," *Karismatika*, vol. 7, no. 1, pp. 10–23, 2021.
- [12] R. Nofitri and N. Irawati, "Integrasi Metode Neive Bayes Dan Software Rapidminer Dalam Analisis Hasil Usaha Perusahaan Dagang," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 6, no. 1, pp. 35–42, 2019, doi: 10.33330/jurteks.v6i1.393.
- [13] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *J. Sist. Komput. dan Inform.*, vol. 1, no. 2, p. 151, 2020, doi: 10.30865/json.v1i2.1955.
- [14] D. Novianti, "Implementasi Algoritma Naive Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 49–54, 2019, doi: 10.31294/p.v21i1.4979.
- [15] Muhammad Romzi and B. Kurniawan, "Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma," *JTIM J. Tek. Inform. Mahakarya*, vol. 03, no. 2, pp. 37–44, 2020.