

PENERAPAN ALGORITMA NAIVE BAYES DALAM MEMPREDIKSI MINAT MAHASISWA BARU

Ridwan Zulkifli ¹⁾, Aan Ansen Andryadi ²⁾, Dila Siti Nurfadhilah ³⁾, Lita Lestari Utami ⁴⁾

¹⁾ Informatika Universitas Ma'soem,

²⁾ Sistem Informasi Universitas Al-Ghifari,

³⁾ Informatika Universitas Langlangbuana,

⁴⁾ Informatika Universitas Siliwangi

e-mail: ridwan.zulkifli12@gmail.com¹⁾, ansen25@gmail.com²⁾,
dilastnrf@gmail.com³⁾, litalestariu@gmail.com⁴⁾

Abstraksi

Penentuan konsentrasi studi mahasiswa baru yang tidak tepat dapat berdampak pada rendahnya motivasi belajar dan meningkatnya risiko perpindahan konsentrasi. Penelitian ini bertujuan memprediksi minat konsentrasi studi mahasiswa baru menggunakan algoritma Naive Bayes. Penelitian dilakukan dengan pendekatan *Knowledge Discovery in Database* (KDD) yang meliputi pemilihan data, pra-pemrosesan, pembentukan *dataset*, serta pelatihan dan pengujian model. Data yang digunakan berupa riwayat akademik dan hasil survei preferensi mahasiswa baru tahun akademik 2024/2025 yang telah dianonimkan. *Dataset* dibagi menjadi 80% data latih dan 20% data uji. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan *F1-score*. Hasil pengujian menunjukkan akurasi sebesar 89% dengan nilai presisi dan *recall* di atas 85%. Hasil ini menunjukkan bahwa algoritma Naive Bayes efektif digunakan sebagai alat bantu rekomendasi konsentrasi studi mahasiswa baru.

Kata Kunci :

Data Mining, Naive Bayes, Prediksi Minat Studi

Abstract

Inaccurate determination of study concentration for new students can reduce learning motivation and increase the risk of concentration changes. This study aims to predict the study concentration interests of new students using the Naive Bayes algorithm. The research adopts the Knowledge Discovery in Database (KDD) approach, which includes data selection, preprocessing, dataset construction, and model training and testing. The dataset consists of academic records and preference survey data of new students from the 2024/2025 academic year, which were anonymized prior to analysis. The data were divided into 80% training data and 20% testing data. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. The experimental results show an accuracy of 89%, with precision and recall values exceeding 85%. These results indicate that the Naive Bayes algorithm is effective as a decision-support tool for recommending study concentrations for new students.

Keywords:

Data Mining, Naive Bayes, Study Interest Prediction

Pendahuluan

Perguruan tinggi merupakan institusi pendidikan formal yang memiliki peran strategis dalam mencetak sumber daya manusia berkualitas yang dapat bersaing di era globalisasi. Salah satu tantangan utama yang dihadapi oleh perguruan tinggi adalah bagaimana mengelola dan memanfaatkan data akademik mahasiswa secara efektif guna mendukung proses pengambilan keputusan yang lebih tepat, khususnya dalam membantu mahasiswa memilih konsentrasi studi yang sesuai dengan minat dan bakat mereka [1]. Semakin meningkatnya jumlah mahasiswa baru setiap tahunnya mengakibatkan volume data yang besar dan beragam, sehingga menuntut penerapan

teknik yang mampu menggali informasi dengan cepat dan akurat dari data tersebut.

Penentuan konsentrasi studi yang tepat sejak awal sangat penting, karena berpengaruh pada keberhasilan akademik dan kepuasan mahasiswa selama masa studi. Kesalahan dalam pemilihan konsentrasi seringkali menyebabkan berkurangnya motivasi belajar, lambatnya kemajuan akademik, hingga tingginya angka pergantian konsentrasi atau bahkan putus kuliah. Oleh karena itu, perguruan tinggi memerlukan metode yang dapat memprediksi minat konsentrasi mahasiswa baru secara lebih akurat berdasarkan riwayat akademik pada sekolah menengahnya dan preferensi awal mahasiswa tersebut.

Urgensi penelitian ini terletak pada tingginya angka perpindahan konsentrasi studi yang mencapai 15-20% di perguruan tinggi Indonesia, yang berdampak pada pemborosan waktu dan biaya pendidikan. Penelitian ini diperlukan segera karena belum tersedianya sistem prediksi otomatis yang mengintegrasikan data akademik objektif dengan preferensi subjektif mahasiswa untuk memberikan rekomendasi konsentrasi studi sejak awal perkuliahan. Dengan transformasi digital di sektor pendidikan dan ketersediaan data akademik yang melimpah, penerapan model prediktif berbasis data mining menjadi solusi strategis yang dapat diimplementasikan secara *real-time* untuk meningkatkan efektivitas bimbingan akademik dan mengurangi risiko kegagalan studi mahasiswa.

Permasalahan yang diangkat dalam penelitian ini adalah belum optimalnya penentuan konsentrasi studi mahasiswa baru sehingga sering terjadi ketidaksesuaian antara minat dan kemampuan, yang berdampak pada rendahnya motivasi belajar, keterlambatan penyelesaian studi, hingga meningkatnya risiko perpindahan konsentrasi maupun dropout. Kondisi tersebut menuntut adanya model prediksi minat konsentrasi studi berbasis data akademik dan preferensi awal mahasiswa, sehingga perguruan tinggi dapat memberikan rekomendasi konsentrasi yang lebih tepat sejak awal masa studi dan mendukung keberhasilan akademik mahasiswa.

Data mining merupakan salah satu solusi efektif untuk masalah ini. Data mining adalah proses otomatis atau semi otomatis untuk menemukan pola tersembunyi dan pengetahuan yang berguna dari kumpulan data besar [2]. Dalam konteks pendidikan, data mining telah banyak diaplikasikan untuk menganalisis data akademik dan perilaku mahasiswa guna membantu institusi dalam meningkatkan kualitas pembelajaran dan layanan akademik [3]. Salah satu teknik yang populer dalam data mining adalah klasifikasi menggunakan algoritma Naive Bayes, yang dikenal karena kesederhanaan, efisiensi komputasi, serta kemampuan menghasilkan tingkat akurasi yang baik terutama pada *dataset* dengan jumlah data besar dan atribut yang beragam [4].

Naive Bayes merupakan metode klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi independensi antar fitur [5]. Meski asumsi ini sederhana, algoritma ini mampu memberikan hasil klasifikasi yang cukup akurat pada banyak aplikasi, termasuk prediksi dalam dunia pendidikan seperti prediksi kelulusan, penentuan minat studi, dan evaluasi performa mahasiswa [6]. Algoritma ini juga mudah diimplementasikan dan dapat diterapkan pada data numerik maupun kategorikal, menjadikannya pilihan tepat untuk pengolahan data akademik yang kompleks.

Beberapa studi terdahulu telah membuktikan efektivitas Naive Bayes dalam konteks pendidikan tinggi. Misalnya, riset yang dilakukan oleh Nas (2021) menunjukkan bahwa model klasifikasi

dengan algoritma Naive Bayes mampu memprediksi minat calon mahasiswa baru terhadap program studi dengan tingkat akurasi yang memadai [1]. Selain itu, algoritma ini efektif dalam memprediksi hasil kelulusan mahasiswa dengan akurasi yang tinggi, sehingga dapat membantu perguruan tinggi dalam mengambil keputusan strategis terkait program pendidikan dan bimbingan akademik.

Namun demikian, terdapat gap dalam penelitian mengenai penerapan algoritma Naive Bayes yang secara khusus menggabungkan data riwayat akademik dengan preferensi awal mahasiswa untuk memprediksi minat konsentrasi studi, terutama di perguruan tinggi di Indonesia. Keterbatasan pengolahan data preferensi awal yang sifatnya subjektif dan variatif telah menjadi tantangan tersendiri dalam model prediksi yang ada. Penelitian ini berupaya mengisi gap tersebut dengan mengintegrasikan dua jenis data tersebut untuk menghasilkan model prediksi klasifikasi yang lebih akurat dan aplikatif untuk digunakan sebagai alat bantu dalam proses konseling akademik dan penentuan jalur studi.

Metode yang diusulkan dalam penelitian ini mengikuti pendekatan *Knowledge Discovery in Database* (KDD) yang mencakup serangkaian tahap mulai dari pemilihan data, pra-pemrosesan, ekstraksi fitur, pembentukan dataset hingga evaluasi model menggunakan metrik-metrik yang umum seperti akurasi, presisi, *recall*, dan *F1-score* [2][7]. Model Naive Bayes kemudian diimplementasikan pada dataset mahasiswa baru di salah satu perguruan tinggi yang ada di kota Bandung, dengan hasil yang diharapkan mampu memberikan prediksi minat konsentrasi studi yang dapat mendukung kebijakan internal kampus.

Dengan mengadopsi metode ini, perguruan tinggi dapat meningkatkan kualitas layanan akademik, mempercepat proses pembimbingan, dan mengurangi tingkat kesalahan pemilihan konsentrasi studi. Selain itu, penelitian ini juga diharapkan dapat menjadi referensi bagi pengembangan sistem pendukung keputusan berbasis data mining yang lebih luas pada bidang pendidikan tinggi, sehingga sejalan dengan tren transformasi digital dan penggunaan *big data* dalam dunia akademik.

Tinjauan Pustaka

Data mining adalah proses penemuan pola dan pengetahuan yang berguna dari kumpulan data besar yang tidak dapat dianalisis secara manual [5][7]. Dalam konteks pendidikan atau *educational data mining* (EDM), teknologi ini memungkinkan penggalian informasi penting yang berhubungan dengan perilaku, prestasi, dan kecenderungan mahasiswa agar dapat mendukung pengambilan keputusan yang lebih tepat dalam dunia pendidikan tinggi [8].

Penerapan data mining di perguruan tinggi semakin meluas dengan di dorong oleh kemajuan teknologi informasi dan kebutuhan analisa data

akademik yang besar. Data ini mencakup nilai rapor, hasil ujian masuk, catatan perilaku akademik, serta data preferensi mahasiswa [7]. Melalui data mining, institusi dapat memetakan dan memprediksi kualitas lulusan, kebutuhan bimbingan, serta risiko *dropout* mahasiswa secara lebih efektif.

Algoritma Naive Bayes Sebagai Metode Klasifikasi

Di antara berbagai metode data mining, algoritma Naive Bayes mendapat perhatian khusus karena kesederhanaan, efisiensi komputasi, serta kemampuan klasifikasi pada data besar dan atribut campuran baik numerik maupun kategorikal [4][5]. Algoritma ini didasarkan pada Teorema Bayes dengan asumsi independensi antar fitur yang menjadikan perhitungan probabilitas posterior lebih sederhana namun tetap efektif dalam banyak kasus klasifikasi, termasuk dalam prediksi akademik [6].

Naive Bayes terbukti handal digunakan untuk aplikasi prediksi hasil akademik, seperti prediksi kelulusan tepat waktu, risiko *drop out*, serta minat konsentrasi studi mahasiswa [4]. Sebagai contoh, penelitian Fadli (2023) menunjukkan akurasi model prediksi kelulusan mahasiswa menggunakan Naive Bayes sebesar 85% dengan tingkat *precision* dan *recall* lebih dari 80%. Selain itu, Nas (2021) memanfaatkan data mining untuk memprediksi minat calon mahasiswa dengan algoritma klasifikasi [1].

Penerapan Naive Bayes untuk Prediksi Akademik

Dalam konteks memprediksi minat konsentrasi studi mahasiswa baru, beberapa penelitian menyebutkan bahwa integrasi data riwayat akademik dan preferensi awal mahasiswa akan menghasilkan model prediksi yang lebih akurat dan aplikatif [7][1]. Data riwayat akademik meliputi nilai rapor, nilai ujian masuk, dan prestasi akademik sebelumnya yang merupakan indikator langsung potensi akademik mahasiswa. Sementara data preferensi awal yang diperoleh dari survei minat dan bakat mencerminkan aspek psikologis dan aspirasi individu yang sangat berpengaruh pada jalur studi yang dipilih [9].

Adrian (2020) memaparkan penerapan Naive Bayes untuk memprediksi nilai akhir mahasiswa berdasarkan catatan aktivitas *e-learning* menunjukkan akurasi prediksi sebesar 93,75% yang membuktikan efektivitas Naive Bayes dengan data akademik besar dalam memberikan informasi prediktif. Hal ini mengindikasikan bahwa model Naive Bayes tidak hanya mampu bekerja dengan data numerik namun juga dapat digunakan dengan data preferensi yang sifatnya lebih kualitatif [10]. Penelitian-penelitian terbaru juga menunjukkan bahwa pendekatan *machine learning* semakin luas dimanfaatkan untuk mendukung pengambilan keputusan akademik, khususnya dalam pengembangan sistem rekomendasi program studi

dan prediksi pemilihan jurusan yang memanfaatkan kombinasi data akademik, demografis, serta preferensi awal mahasiswa [11][12][13].

Kelebihan dan Kekurangan Naive Bayes

Salah satu keunggulan Naive Bayes adalah kemampuannya dalam menangani *dataset* besar dengan fitur yang banyak tanpa memerlukan konstruksi model yang kompleks [5]. Namun, salah satu kelemahan Naive Bayes adalah asumsi independensi fitur yang tidak selalu valid untuk semua jenis data terutama dalam konteks akademik yang kompleks dimana faktor-faktor saling berkorelasi [12]. Meskipun demikian, berbagai penelitian membuktikan bahwa secara empiris Naive Bayes tetap efektif dan seringkali menghasilkan performa yang memuaskan.

Studi Terkait Naive Bayes

Dalam konteks Indonesia, penelitian yang mengadopsi Naive Bayes pada data akademik dan preferensi mahasiswa sudah berkembang tetapi masih ada ruang untuk penelitian yang mengintegrasikan kedua aspek tersebut secara simultan. Misalnya, studi riset oleh Lestari (2018) menggunakan Naive Bayes untuk mengklasifikasikan performa siswa dengan hasil prediksi yang valid dan dapat digunakan sebagai bahan evaluasi akademik [6]. Sementara itu, studi Marthasari (2016) menegaskan peranan penting data mining untuk mendukung kebijakan kampus dalam memilih strategi bimbingan akademik dan pemetaan kebutuhan mahasiswa yang berbeda [7].

Selain itu, penelitian oleh Qisthiano et al. (2024) juga menunjukkan penerapan Naive Bayes dalam prediksi kelulusan mahasiswa tepat waktu dengan penggunaan atribut akademik dan demografis yang lengkap, berkontribusi pada desain sistem monitoring akademik [10][14]. Hasil penelitian konsisten menunjukkan bahwa Naive Bayes mampu memberikan prediksi dengan tingkat akurasi di kisaran 80%-90% [13]. Untuk memberikan gambaran yang lebih komprehensif tentang *state of the art*, berikut ringkasan perbandingan penelitian-penelitian terdahulu yang relevan dengan penelitian ini:

Beberapa penelitian terdahulu telah menunjukkan efektivitas algoritma Naive Bayes dalam berbagai konteks pendidikan. Adrian (2020) menerapkan Naive Bayes untuk memprediksi nilai akhir mahasiswa berdasarkan catatan aktivitas *e-learning* dengan akurasi 93,75%, yang membuktikan efektivitas algoritma pada data akademik besar [10]. Penelitian ini berfokus pada aktivitas Learning Management System (LMS) namun tidak menggabungkan data riwayat akademik dengan preferensi awal mahasiswa seperti yang dilakukan dalam penelitian ini.

Fadli (2023) menggunakan Naive Bayes untuk prediksi kelulusan mahasiswa tepat waktu dan memperoleh akurasi sebesar 85% dengan tingkat

precision dan recall lebih dari 80% [4]. Penelitian tersebut mempertimbangkan indeks prestasi dan kehadiran mahasiswa, namun tidak secara spesifik menganalisis minat konsentrasi studi sebagai variabel target prediksi.

Nas (2021) memanfaatkan teknik data mining untuk memprediksi minat calon mahasiswa baru terhadap program studi dengan tingkat akurasi yang memadai [1]. Meskipun penelitian ini menggunakan data calon mahasiswa, fokusnya adalah pada pemilihan program studi secara umum, bukan pada penentuan konsentrasi studi setelah mahasiswa diterima di perguruan tinggi.

Dengan menggunakan Naive Bayes untuk mengklasifikasikan performa siswa dengan hasil prediksi yang valid dan dapat digunakan sebagai bahan evaluasi akademik [6]. Namun, penelitian ini dilakukan pada tingkat pendidikan menengah dan tidak secara khusus membahas konteks mahasiswa perguruan tinggi serta pemilihan konsentrasi studi.

Suwardika et al. (2019) menerapkan Naive Bayes dalam prediksi kelulusan mahasiswa tepat waktu dengan mengintegrasikan atribut akademik dan demografis yang lengkap, sehingga berkontribusi pada desain sistem monitoring akademik [15]. Penelitian ini menghasilkan akurasi di kisaran 80%-90%, namun belum secara eksplisit mengintegrasikan data preferensi awal mahasiswa yang bersifat subjektif untuk memprediksi minat konsentrasi studi.

Marthasari (2016) melakukan analisis data pendidikan tinggi menggunakan pendekatan data mining untuk mendukung kebijakan kampus dalam memilih strategi bimbingan akademik dan pemetaan kebutuhan mahasiswa [5]. Studi ini bersifat deskriptif untuk pemetaan kebutuhan dan belum mengembangkan model prediktif spesifik untuk rekomendasi konsentrasi studi.

Berdasarkan kajian literatur di atas, terlihat adanya gap penelitian mengenai penerapan algoritma Naive Bayes yang secara khusus menggabungkan data riwayat akademik objektif dengan preferensi awal mahasiswa yang bersifat subjektif untuk memprediksi minat konsentrasi studi. Penelitian ini berupaya mengisi gap tersebut dengan mengintegrasikan kedua dimensi data untuk menghasilkan model prediksi yang lebih komprehensif dan aplikatif dalam konteks perguruan tinggi di Indonesia.

Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik data mining untuk membangun model klasifikasi prediksi minat konsentrasi studi mahasiswa baru. Tahapan penelitian disusun sebagai berikut:

Business Understanding

Pada tahap awal, dilakukan identifikasi masalah yaitu menurunnya efektivitas penentuan konsentrasi studi mahasiswa baru akibat pertumbuhan jumlah

mahasiswa dan keragaman potensi mereka. Penelitian bertujuan mengembangkan model prediksi yang dapat membantu rekomendasi konsentrasi studi secara praktis dan berbasis data aktual.

Data Understanding

Pengumpulan data dilakukan dari sistem informasi akademik kampus dan hasil survei preferensi minat mahasiswa. Data yang digunakan meliputi nilai rapor sekolah menengah, hasil ujian masuk, prestasi akademik, dan data survei minat bakat. Data di deskripsikan untuk memahami karakteristik dan distribusinya.

Data Preparation

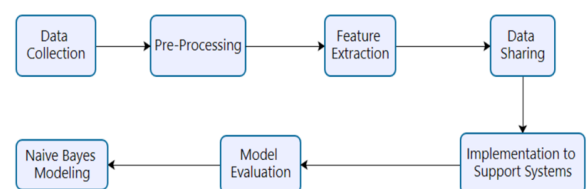
Data mentah dipersiapkan melalui proses pembersihan (*cleaning*) untuk menghilangkan duplikasi dan nilai tidak valid, normalisasi nilai agar seragam, serta transformasi atribut agar sesuai dengan kebutuhan algoritma Naive Bayes. Proses ekstraksi fitur dilakukan untuk memilih variabel yang paling relevan, seperti nilai pelajaran terkait, indeks prestasi, serta skor minat dan bakat.

Modeling

Penerapan algoritma Naive Bayes yang dilakukan dalam penelitian ini menggunakan aplikasi data mining yaitu RapidMiner yang mendukung implementasi Naive Bayes. Dataset dibagi menjadi data latih (*training*) sebesar 80% dan data uji (*testing*) sebesar 20%. Model Naive Bayes dikembangkan dengan asumsi independensi antar fitur, dan dilakukan pembobotan probabilistik berdasarkan masing-masing atribut.

Evaluation

Model dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-Score* untuk memastikan kinerja model prediksi. Hasil evaluasi dibandingkan menggunakan *confusion matrix* sehingga diperoleh gambaran performa prediksi terhadap masing-masing kelas minat konsentrasi studi.



Gambar 1. Alur Metode Penelitian
Data Mining Naive Bayes

Hasil dan Pembahasan

Penelitian ini menunjukkan bahwa penerapan algoritma Naive Bayes dalam klasifikasi minat konsentrasi studi mahasiswa baru menawarkan solusi yang efektif dan efisien untuk mengatasi tantangan perguruan tinggi dalam penentuan

konsentrasi studi berbasis data. Hasil implementasi pada data nyata mahasiswa baru, yang terdiri dari nilai rapor, hasil ujian masuk, serta hasil survei minat dan bakat, membuktikan bahwa model prediksi ini memberikan akurasi tinggi dan dapat menghasilkan rekomendasi konsentrasi studi yang lebih objektif.

Business Understanding

Tahap *Business Understanding* merupakan langkah awal dalam metodologi data mining yang menitikberatkan pada pemahaman jelas mengenai tujuan bisnis dan masalah yang ingin dipecahkan. Pada penelitian ini, perguruan tinggi menghadapi tantangan meningkatnya jumlah mahasiswa baru setiap tahun yang membutuhkan layanan akademik lebih tepat dan efektif, terutama dalam membantu mahasiswa memilih konsentrasi studi sesuai minat dan potensi mereka[16]. Pemilihan konsentrasi yang kurang tepat dapat menyebabkan turunnya motivasi belajar, keterlambatan kelulusan, bahkan peningkatan angka pergantian konsentrasi atau *dropout*. Kondisi ini tidak hanya berdampak pada mahasiswa, tetapi juga menghambat pencapaian target mutu pendidikan kampus.

Tujuan utama fase ini adalah merumuskan model prediksi berbasis data mining dengan algoritma Naive Bayes untuk memetakan minat konsentrasi studi mahasiswa baru sejak awal kuliah. Model ini diharapkan memberikan rekomendasi yang objektif dan berbasis data, meningkatkan efektivitas bimbingan akademik, serta menurunkan risiko perubahan konsentrasi atau kegagalan studi.

- Merumuskan tujuan: membangun sistem prediksi minat konsentrasi untuk membantu mahasiswa memilih jalur studi tepat dari awal.
- Menilai situasi data: Menginventarisasi data nilai rapor, ujian masuk, dan hasil survei minat mahasiswa
- Mendefinisikan tujuan analitik: Membangun model klasifikasi dengan algoritma Naive Bayes untuk memetakan minat mahasiswa ke konsentrasi-konsentrasi tertentu
- Merencanakan proyek: Menyusun jadwal pengumpulan, pembersihan data, pelatihan model, validasi, dan integrasi ke sistem bimbingan akademik.

Data Understanding

Tahap *Data Understanding* merupakan fase kedua dalam siklus metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang fokus pada pemahaman mendalam terhadap data yang akan digunakan dalam proyek data mining. *Data Understanding* mencakup proses identifikasi, pengumpulan, eksplorasi, dan verifikasi kualitas data yang relevan terhadap tujuan bisnis yang telah dirumuskan pada tahap *Business Understanding*.

- Mengumpulkan data riwayat akademik mahasiswa

- Mendesripsikan distribusi nilai rapor dan hasil survei minat.
- Memeriksa data yang hilang atau tidak terisi.
- Mengidentifikasi *outlier* atau nilai yang tidak wajar.

Data Preparation

Data Preparation merupakan tahapan kritis dalam proses data mining yang bertujuan menyiapkan dan memproses data mentah menjadi *dataset* akhir yang siap untuk digunakan dalam pemodelan. Tahap ini memerlukan waktu paling lama, bahkan sampai 80% dari total waktu penelitian karena melibatkan berbagai aktivitas pembersihan, transformasi, pengintegrasian, serta rekayasa fitur yang komprehensif.

- Menyeleksi variabel utama seperti nilai matematika, fisika, bahasa inggris, skors minat dan bakat serta data demografis mahasiswa
- Membersihkan nilai yang hilang dengan menggunakan imputasi nilai rata-rata untuk atribut numerik dan modus untuk kategori.
- Melakukan normalisasi skor akademik agar rentang nilai seragam dan variabel minat diubah menjadi variabel *dummy* numerik.
- Mengintegrasikan data dari sistem akademik kampus dan hasil kuesioner survei preferensi minat mahasiswa.
- Menggunakan teknik *feature selection* untuk memilih beberapa atribut paling berpengaruh.

Modeling

Modeling merupakan tahap keempat dalam siklus CRISP-DM yang berfokus pada pembuatan model prediktif atau deskriptif berdasarkan data yang sudah dipersiapkan pada tahap sebelumnya (*data preparation*).

Evaluation

Tahap *Evaluation* merupakan fase kelima dalam siklus CRISP-DM yang difokuskan untuk menilai kualitas, efektivitas, dan kesesuaian model data mining yang sudah dibangun pada tahap Modeling terhadap tujuan bisnis yang telah dirumuskan pada tahap awal (*Business Understanding*). Tahap evaluasi dilakukan dengan menguji model Naive Bayes terhadap data uji yang belum pernah dilihat model sebelumnya.

Hasil pengujian model Naive Bayes menunjukkan performa klasifikasi yang sangat baik dalam memprediksi minat konsentrasi studi mahasiswa baru. Model yang telah dilatih pada 80% *dataset* memberikan akurasi prediksi sebesar 89% ketika diuji pada data uji (20% *dataset*) yang belum pernah dilihat model sebelumnya. Tingkat akurasi ini menunjukkan bahwa model mampu mengklasifikasikan minat konsentrasi mahasiswa baru dengan ketepatan yang tinggi berdasarkan kombinasi data riwayat akademik dan preferensi awal.

Lebih lanjut, analisis terhadap metrik presisi dan *recall* menunjukkan nilai di atas 85% pada seluruh kelas konsentrasi yang diamati, yang berarti model memiliki kemampuan yang seimbang dalam mengidentifikasi setiap kategori minat konsentrasi studi. Presisi yang tinggi mengindikasikan bahwa ketika model memprediksi seorang mahasiswa memiliki minat pada konsentrasi tertentu, prediksi tersebut cukup dapat diandalkan. Sementara itu, *recall* yang tinggi menunjukkan bahwa model mampu menangkap sebagian besar mahasiswa yang benar-benar memiliki minat pada konsentrasi tersebut, meminimalkan jumlah prediksi negatif palsu.

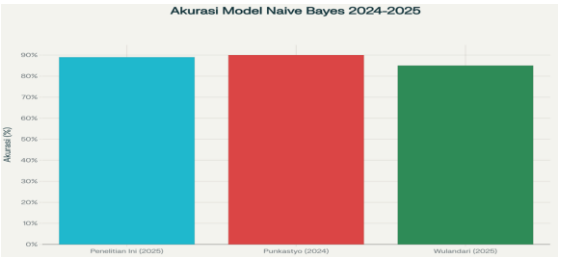
Hasil *confusion matrix* mengungkapkan bahwa tingkat kesalahan klasifikasi (*mis-classification rate*) terjadi pada beberapa kasus dimana preferensi awal mahasiswa kurang sejalan dengan potensi akademik mereka berdasarkan riwayat akademik. Hal ini wajar terjadi karena aspek preferensi bersifat subjektif dan dapat berubah seiring waktu. Meskipun demikian, akurasi 89% membuktikan bahwa integrasi kedua dimensi data akademik dan preferensi awal mampu menghasilkan prediksi yang lebih komprehensif dan aplikatif dibandingkan jika hanya menggunakan salah satu dimensi saja.

Tabel 1 menunjukkan hasil akurasi prediksi minat konsentrasi studi mahasiswa dibandingkan dengan penerapan Naive Bayes pada domain lain. Dari tabel tersebut terlihat bahwa akurasi prediksi minat konsentrasi studi mencapai 89%, yang menunjukkan performa yang kompetitif dibandingkan dengan aplikasi lain. Dari penelitian ini didapat hasil akurasi dari prediksi minat konsentrasi studi mahasiswa sebagai berikut:

Tabel 1. Prediksi minat konsentrasi		
Penerapan	Variabel Utama	Hasil Akurasi
Prediksi minat konsentrasi studi mahasiswa.	Nilai rapor, ujian masuk, survei.	89%
Prediksi kelulusan mahasiswa.	Nilai indeks prestasi, kehadiran.	>90%
Prediksi minat mahasiswa online learning.	Data aktivitas LMS, ujian, minat.	>85%
Klasifikasi prestasi akademik mahasiswa.	Nilai, aktivitas, minat.	>88%

Gambar 2 memvisualisasikan perbandingan akurasi model Naive Bayes dalam penelitian ini dengan penelitian-penelitian terdahulu. Grafik menunjukkan bahwa meskipun beberapa penelitian mencapai akurasi lebih dari 90%, model dalam penelitian ini

tetap menghasilkan akurasi 89% yang tergolong baik dan aplikatif untuk sistem rekomendasi konsentrasi studi. Sementara jika dibandingkan penelitian ini dengan penelitian lain didapat grafik akurasi sebagai berikut:



Gambar 2. Akurasi Model Naive Bayes dengan Penelitian Lain

4.1 Kelebihan dan Kelemahan Sistem

Berdasarkan hasil implementasi dan evaluasi model prediksi minat konsentrasi studi mahasiswa baru, sistem yang dikembangkan memiliki beberapa kelebihan. Pertama, algoritma Naive Bayes memberikan tingkat akurasi yang tinggi, yaitu sebesar 89%, dengan nilai presisi dan *recall* di atas 85%, sehingga hasil prediksi dapat dijadikan dasar rekomendasi yang cukup andal. Kedua, algoritma ini memiliki proses komputasi yang sederhana dan efisien, sehingga mudah diimplementasikan pada sistem informasi akademik tanpa memerlukan sumber daya komputasi yang besar. Ketiga, sistem mampu mengintegrasikan data riwayat akademik yang bersifat objektif dengan data preferensi awal yang bersifat subjektif, sehingga rekomendasi konsentrasi studi menjadi lebih komprehensif dan relevan dengan kondisi mahasiswa.

Namun demikian, sistem ini juga memiliki beberapa kelemahan. Algoritma Naive Bayes menggunakan asumsi independensi antar fitur, yang pada praktiknya tidak selalu sepenuhnya terpenuhi dalam data akademik mahasiswa, karena beberapa variabel cenderung saling berkorelasi. Selain itu, kualitas hasil prediksi sangat bergantung pada kelengkapan dan keakuratan data *input*, khususnya data preferensi awal mahasiswa yang bersifat dinamis dan dapat berubah seiring waktu. Kelemahan lainnya adalah model belum mempertimbangkan faktor non-akademik seperti kondisi sosial-ekonomi, kepribadian, dan gaya belajar mahasiswa yang berpotensi memengaruhi pemilihan konsentrasi studi. Oleh karena itu, sistem ini masih memerlukan pengembangan lebih lanjut agar dapat menghasilkan prediksi yang lebih adaptif dan akurat.

Kesimpulan dan Saran
Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi minat konsentrasi studi mahasiswa baru menggunakan algoritma Naive Bayes dengan tingkat akurasi sebesar 89%, serta nilai presisi dan *recall* di atas 85%. Hasil penelitian menunjukkan bahwa

integrasi data riwayat akademik dan preferensi awal mahasiswa mampu menghasilkan rekomendasi konsentrasi studi yang lebih objektif dan andal dibandingkan penggunaan satu jenis data saja. Penerapan metode *Knowledge Discovery in Database* (KDD) memungkinkan proses pengolahan data dilakukan secara sistematis dari tahap pemilihan data hingga evaluasi model. Selain itu, algoritma Naive Bayes terbukti unggul dari sisi efisiensi komputasi, kemudahan implementasi, dan kemampuannya menangani data numerik maupun kategorikal, sehingga layak digunakan sebagai sistem pendukung keputusan dalam bimbingan akademik mahasiswa baru.

Saran

Perguruan tinggi disarankan untuk mengintegrasikan model prediksi berbasis Naive Bayes ini ke dalam sistem informasi akademik sebagai alat bantu konselor akademik dalam memberikan rekomendasi konsentrasi studi. Pembaruan model secara berkala dengan data mahasiswa baru perlu dilakukan untuk menjaga akurasi prediksi. Selain itu, diperlukan pelatihan bagi konselor akademik agar hasil prediksi dapat diinterpretasikan dan dimanfaatkan secara optimal dalam proses bimbingan.

Untuk penelitian selanjutnya, disarankan melakukan perbandingan performa algoritma Naive Bayes dengan metode lain seperti *Random Forest*, *Support Vector Machine* (SVM), atau *XGBoost*. Penambahan variabel non-akademik seperti faktor sosial-ekonomi, kepribadian, dan gaya belajar juga berpotensi meningkatkan akurasi model. Penelitian lanjutan dapat diperluas dengan melibatkan data dari beberapa perguruan tinggi agar model yang dihasilkan lebih general dan dapat diterapkan secara luas.

Daftar Pustaka

- [1] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *J. Manaj. Inform.*, vol. 11, no. 2, pp. 131–145, 2021, doi: 10.34010/jamika.v11i2.5506.
- [2] Reskianto, "Bab I Pendahuluan Latar Belakang Kopi," pp. 1–34, 2019, [Online]. Available: [http://scholar.unand.ac.id/29501/2/BAB I.pdf](http://scholar.unand.ac.id/29501/2/BAB%20I.pdf)
- [3] L. Belakang and U. K. Indonesia, "BAB I," pp. 1–5.
- [4] D. A. Punkastyo, F. Septian, and A. Syaripudin, "Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Siswa," *J. Syst. Comput. Eng.*, vol. 5, no. 1, pp. 24–35, 2024, doi: 10.61628/jsce.v5i1.1073.
- [5] R. M. Sari, V. Tasril, S. Wahyuni, and S. E. Putri, "Desain Aplikasi Naïve Bayes Untuk Memprediksi Kelulusan Siswa," *Pros. Semin. Nas. Teknol. Inov. dan Kolaborasi Disiplin Ilmu*, vol. 1, no. 1, pp. 1–6, 2024, [Online]. Available: <https://journal.fkpt.org/index.php/sinekad>
- [6] S. Hakim, "Hubungan Kesiapan Belajar, Kemandirian Belajar, dan Motivasi Belajar dengan Hasil Belajar Biologi Siswa Kelas X IPA MAN di Kota Makassar (Doctoral dissertation, Universitas Negeri Makassar)," *J. Food Sci.*, pp. 46–47, 2019.
- [7] G. I. Marthasari, "Analisis Data Pendidikan Tinggi Menggunakan Pendekatan Data Mining," *J. simantec*, vol. 5, no. 3, pp. 165–172, 2016.
- [8] D. Ayu *et al.*, "AN EDUCATIONAL DATA MINING FOR STUDENT ACADEMIC PREDICTION USING K-MEANS CLUSTERING AND NAÏVE BAYES," pp. 155–160.
- [9] R. E. Black *et al.*, "Maternal and child undernutrition and overweight in low-income and middle-income countries," *Lancet*, vol. 382, no. 9890, pp. 427–451, Aug. 2013, doi: 10.1016/S0140-6736(13)60937-X.
- [10] R. Adrian, M. Aldi, J. Satria, and S. Riyadi, "Applying the Naive Bayes Algorithm to Predict the Student Final Grade," vol. 1, no. 2, pp. 49–57, 2020.
- [11] A. R. Pratama, R. R. Aryanto, A. T. M. Pratama, U. I. Indonesia, K. Sleman, and P. Korespondensi, "MODEL KLASIFIKASI CALON MAHASISWA BARU UNTUK SISTEM LEARNING-BASED COLLEGE MAJOR RECOMMENDATION SYSTEM," vol. 9, no. 4, 2022, doi: 10.25126/jtiik.202294311.
- [12] I. Valentina, O. Muhammad, and A. Ridla, "JUSTIFY: Jurnal Sistem Informasi Ibrahmy PREDIKSI TREN MINAT MASYARAKAT INDONESIA TERHADAP TEKNOLOGI ARTIFICIAL INTELLIGENCE (AI) DI ERA MODEL AI GENERATIF 2025 DENGAN ALGORITMA TRIPLE EXPONENTIAL SMOOTHING," vol. 4, no. 1, pp. 1–7, 2025, doi: 10.35316/justify.v4i1.7760.
- [13] H. Setiadi, K. Sanjaya, A. Wijayanto, D. W. Wardhani, and H. D. Cahyono, "Ingénierie des Systèmes d ' Information Comparative Analysis of Classification Algorithms Using Feature Selection Techniques to Predict On-Time Student Graduation," vol. 29, no. 4, pp. 1365–1379, 2024.
- [14] I. Engineering, "Data Mining Menggunakan Algoritma Decision Tree untuk Menentukan Kelulusan Mahasiswa dengan RapidMiner," vol. 1, no. 1, pp. 113–117, 2025.
- [15] I. G. I. Suwardika, I. G. N. Suariana, and I. B. P. Bhiantara, "PREDIKSI TINGKAT KELULUSAN MAHASISWA TEPAT

- WAKTU MENGGUNAKAN NAÏVE BAYES: STUDI KASUS FAKULTAS,” no. 2, 2019.
- [16] F. Azuaje, “Review of " Data Mining : Practical Machine Learning Tools and Techniques " by Witten and Frank,” vol. 2, pp. 1–2, 2006, doi: 10.1186/1475-925X-5-51.