

BAB I

PENDAHULUAN

Bab ini akan membahas latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, dan sistematika penulisan penelitian ini.

1.1 Latar Belakang

Tingkat kelulusan mahasiswa adalah salah satu parameter utama yang digunakan untuk mengukur kinerja institusi pendidikan tinggi. Di banyak negara, termasuk Indonesia, kelulusan tepat waktu menjadi salah satu indikator penting dalam proses akreditasi program studi serta evaluasi kualitas pendidikan. Data dari Direktorat Jenderal Pendidikan Tinggi menunjukkan bahwa hanya sekitar 52% mahasiswa Indonesia yang berhasil lulus tepat waktu dalam lima tahun terakhir [1]. Angka ini mencerminkan adanya tantangan besar dalam meningkatkan efektivitas sistem pendidikan, mengingat kelulusan tepat waktu memiliki implikasi signifikan terhadap kualitas dan reputasi perguruan tinggi.

Masalah rendahnya tingkat kelulusan tepat waktu tidak hanya terkait dengan faktor internal mahasiswa seperti kemampuan akademik, motivasi, dan kedisiplinan dalam belajar, tetapi juga dipengaruhi oleh faktor eksternal seperti kondisi ekonomi, lingkungan keluarga, dan kebijakan kampus. Banyak mahasiswa yang terhambat kelulusannya akibat ketidakseimbangan antara kemampuan mereka dengan tuntutan akademik, serta faktor sosial-ekonomi yang membatasi akses mereka untuk menyelesaikan pendidikan tepat waktu. Ketidakmampuan untuk mengidentifikasi mahasiswa yang berisiko tinggi gagal lulus tepat waktu secara dini menyebabkan institusi pendidikan kesulitan untuk memberikan dukungan atau intervensi yang dibutuhkan sebelum terlambat.

Salah satu cara yang dapat digunakan untuk memitigasi masalah ini adalah melalui prediksi kelulusan mahasiswa. Dengan menerapkan metode prediktif berbasis data, perguruan tinggi dapat melakukan pemetaan risiko secara lebih awal dan mendalam, sehingga intervensi akademik dapat dilakukan dengan lebih tepat dan efektif. Pendekatan ini memungkinkan pemangku kepentingan untuk

mengidentifikasi mahasiswa yang berisiko tinggi dan memberikan mereka dukungan yang diperlukan agar dapat lulus tepat waktu. Hal ini juga memberikan kesempatan bagi kampus untuk merancang kebijakan yang lebih responsif terhadap kebutuhan mahasiswa [2].

Prediksi kelulusan mahasiswa dapat dilakukan melalui berbagai teknik analisis data, salah satunya adalah klasifikasi, yang merupakan bagian dari bidang Educational Data Mining (EDM). Klasifikasi memanfaatkan data historis mahasiswa untuk memprediksi apakah seorang mahasiswa akan lulus tepat waktu atau tidak. Teknik ini sangat bermanfaat untuk menganalisis data besar dan menemukan pola yang tersembunyi, yang tidak mudah dilihat dengan metode manual [3]. Klasifikasi dapat digunakan untuk memetakan berbagai faktor yang mempengaruhi kelulusan dan membuat prediksi yang lebih akurat mengenai status kelulusan mahasiswa.

Penggunaan ilmu data dan teknik machine learning kini semakin relevan untuk menyelesaikan tantangan ini. Beberapa metode klasifikasi populer seperti Decision Tree, Random Forest, dan Naïve Bayes telah banyak diterapkan. Metode Naïve Bayes dikenal secara khusus karena kemudahannya serta efektivitasnya dalam klasifikasi berbasis probabilistik. Menurut studi [4], Naïve Bayes menunjukkan performa kompetitif di lingkungan akademik, bahkan dengan asumsi independensi fitur yang ketat. Kelebihan metode ini meliputi kecepatan tinggi dalam memproses dataset yang besar dan menghasilkan prediksi dengan tingkat akurasi yang cukup baik. Namun, efektivitas algoritma klasifikasi tidak dapat hanya dinilai berdasarkan akurasi semata. Evaluasi yang lebih mendalam dengan metrik seperti precision, recall, F1-score, dan confusion matrix menjadi sangat penting, terutama dalam kondisi dataset yang tidak seimbang, seperti yang banyak ditemui dalam konteks pendidikan tinggi [5][6].

Implementasi Naïve Bayes dalam penelitian terdahulu memberikan hasil yang cukup bervariasi. Sebagian penelitian mencatat keterbatasan Naïve Bayes dalam menghadapi data yang tidak seimbang tanpa teknik tambahan seperti oversampling dan feature engineering [7][8]. Penelitian [9] melakukan kajian

komparatif antara Naïve Bayes dengan Decision Tree dan menemukan perbedaan signifikan dalam nilai precision dan recall. Hasil penelitian ini menegaskan perlunya evaluasi metrik yang lebih komprehensif dalam penerapan metode klasifikasi pada dataset akademik.

Kesenjangan penelitian terdahulu yang paling mencolok adalah minimnya analisis kuantitatif yang komprehensif mengenai evaluasi metrik performa algoritma klasifikasi. Mayoritas studi sebelumnya cenderung mengandalkan tingkat akurasi saja tanpa memperhatikan dampak imbalance data serta efek fatal dari prediksi false negative dalam dunia pendidikan [10][11].

Penelitian ini bertujuan secara spesifik untuk menutup kesenjangan tersebut, dengan menekankan evaluasi komprehensif terhadap performa algoritma Naïve Bayes dalam klasifikasi kelulusan mahasiswa. Pendekatan kuantitatif dengan data historis mahasiswa diharapkan mampu menghasilkan metrik performa yang lebih lengkap dan memberikan wawasan mengenai faktor yang berpengaruh terhadap akurasi prediksi kelulusan. Penelitian ini juga mempertimbangkan implementasi teknik balancing data serta metode validasi silang, agar model yang dibangun tidak hanya mampu bekerja secara cepat, tetapi juga reliabel serta generalizable dalam berbagai kondisi dataset akademik.

Dari perspektif teoritis, penelitian ini akan memberikan kontribusi penting dengan memperluas pemahaman tentang efektivitas Naïve Bayes di bidang pendidikan tinggi. Evaluasi komprehensif terhadap algoritma ini diharapkan mampu mendorong pengembangan sistem prediksi yang lebih akurat, efisien, dan aplikatif. Secara praktis, hasil penelitian ini akan menjadi landasan bagi institusi pendidikan untuk merumuskan kebijakan akademik yang tepat sasaran. Dengan prediksi kelulusan yang akurat, institusi dapat melakukan intervensi tepat waktu kepada mahasiswa berisiko, meningkatkan efisiensi akademik, serta memastikan mutu lulusan yang tinggi.

Oleh karena itu, tujuan utama penelitian ini adalah mengevaluasi efektivitas Naïve Bayes dalam mengklasifikasikan kelulusan mahasiswa melalui pengukuran metrik performa yang komprehensif. Harapannya, penelitian ini mampu

memberikan kontribusi nyata dalam pengembangan sistem pendukung keputusan akademik berbasis data yang andal dan relevan dengan kebutuhan institusi pendidikan tinggi.

1.2 Perumusan Masalah

Pada bagian perumusan masalah, diuraikan mengenai permasalahan, metode usulan serta pertanyaan pada penelitian ini:

1.2.1 Permasalahan

Tingkat kelulusan mahasiswa yang rendah atau tidak tepat waktu menjadi tantangan serius bagi institusi perguruan tinggi. Permasalahan ini berdampak pada efisiensi akademik, kualitas lulusan, dan akreditasi institusi. Saat ini, belum banyak tersedia sistem prediksi yang objektif dan berbasis data untuk mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu sejak dini. Beberapa penelitian sebelumnya telah menggunakan algoritma klasifikasi untuk memprediksi kelulusan mahasiswa, namun fokus utama hanya pada akurasi tanpa evaluasi menyeluruh, serta terbatasnya penerapan metode Naïve Bayes dengan metrik evaluasi yang lengkap seperti precision, recall, F1-score, dan AUC. Oleh sebab itu, penelitian ini mengusulkan penggunaan metode Naïve Bayes yang sederhana dan efisien untuk membangun model prediktif yang lebih akurat dan terukur dalam konteks manajemen akademik.

1.2.2 Metode Usulan

Pada penelitian ini, akan mengusulkan penggunaan metode Naïve Bayes sebagai teknik klasifikasi berbasis probabilitas untuk memprediksi kelulusan mahasiswa. Dataset yang digunakan terdiri dari berbagai atribut demografis, sosial-ekonomi, dan akademik mahasiswa. Sebelum proses klasifikasi, data akan melalui tahap preprocessing termasuk encoding variabel kategorikal dan normalisasi data numerik agar dapat diolah secara optimal oleh model Naïve Bayes. Model klasifikasi yang dibangun akan dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score untuk menilai efektivitas metode ini dalam mengklasifikasikan status kelulusan mahasiswa.

1.2.3 Pertanyaan Penelitian

Berdasarkan latar belakang dan permasalahan yang telah diuraikan, penelitian ini difokuskan untuk menjawab pertanyaan utama mengenai efektivitas metode Naïve Bayes dalam mengklasifikasikan kelulusan mahasiswa, yaitu:

1. Bagaimana performa metode Naïve Bayes dalam mengklasifikasikan kelulusan mahasiswa dilihat dari metrik evaluasi seperti Accuracy, Precision, Recall, F1 Score, dan Specificity?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Adapun tujuan akhir dari penelitian ini adalah untuk mengembangkan model klasifikasi kelulusan mahasiswa yang efektif menggunakan metode Naïve Bayes :

1. Mengetahui performa metrik Accuracy, Precision, Recall, dan F1-Score yang diperoleh dari proses klasifikasi kelulusan mahasiswa menggunakan metode Naïve Bayes.

1.3.2 Manfaat Penelitian

Dalam Penelitian ini memiliki beberapa manfaat, yakni sebagai berikut:

1. Manfaat Teoritis:
 - Memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya di bidang data mining dan klasifikasi menggunakan metode Naïve Bayes.
 - Menambah referensi dan gambaran efektivitas metode Naïve Bayes dalam prediksi kelulusan mahasiswa.
2. Manfaat Praktis:
 - Membantu perguruan tinggi dalam mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu secara lebih akurat.
 - Memfasilitasi intervensi tepat sasaran untuk meningkatkan tingkat kelulusan mahasiswa.

- Menjadi dasar pengembangan sistem pendukung keputusan berbasis data untuk pengelolaan akademik.
- Memberikan panduan dalam penyusunan kebijakan yang mendukung keberhasilan studi mahasiswa.

1.4 Batasan Penelitian

Adapun batasan masalah yang dihadapi saat melakukan penelitian ini adalah sebagai berikut:

1. Studi kasus yang digunakan adalah prediksi kelulusan mahasiswa pada program sarjana di sebuah institusi pendidikan tinggi.
2. Data yang digunakan berasal dari dataset "Higher Education Predictors of Student Retention" yang diunduh dari situs web *kaggle.com* ([Predict students' dropout and academic success](#)).
3. Dataset mencakup variabel numerik dan kategorikal seperti status pernikahan, jurusan, kehadiran kelas, jumlah satuan kredit, dan nilai akademik.
4. Penelitian ini hanya menggunakan data dari satu semester untuk setiap angkatan mahasiswa sehingga tidak mencakup analisis data dalam rentang waktu yang lebih panjang.
5. Metode penelitian yang digunakan adalah klasifikasi dengan algoritma Naïve Bayes.
6. Output yang dihasilkan berupa klasifikasi status kelulusan mahasiswa, yaitu lulus tepat waktu atau tidak.
7. Evaluasi kinerja model menggunakan metrik Accuracy, Precision, Recall, F1-Score, dan Confusion Matrix.

1.5 Sistematika Proposal

Sistematika proposal tugas akhir ini yaitu terdiri dari lima bab , dari masing-masing bab merupakan rangkaian sistematis dari perumusan masalah yang telah dicantumkan pada BAB I dengan sistematis sebagai berikut:

BAB I PENDAHULUAN

Menjelaskan mengenai latar belakang, perumusan masalah, permasalahan, metode usulan, pertanyaan penelitian, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan terkait penelitian yang dilakukan.

BAB II TINJAUAN PUSTAKA

Bab ini memuat kajian teori yang mendasari penelitian, termasuk penjelasan tentang klasifikasi dalam data mining, algoritma Naïve Bayes, serta metrik evaluasi performa model seperti Accuracy, Precision, Recall, F1 Score, dan Specificity. Selain itu, bab ini menyertakan ulasan terhadap penelitian terdahulu yang relevan untuk menunjukkan posisi penelitian ini dalam konteks literatur yang ada, serta menyusun kerangka berpikir teoritis sebagai dasar analisis.

BAB III METODE PENELITIAN

Bab ini menjelaskan pendekatan dan jenis penelitian yang digunakan, sumber dan jenis data, teknik pengumpulan data, serta metode atau alat analisis data. Penjabaran mencakup bagaimana model klasifikasi Naïve Bayes akan dibangun dan diuji, serta teknik evaluasi yang digunakan untuk mengukur performa model terhadap data kelulusan mahasiswa.

BAB IV PEMBAHASAN

Bab ini menyajikan hasil implementasi model prediksi kelulusan menggunakan algoritma Naïve Bayes. Penyajian hasil dilakukan dalam bentuk data statistik, visualisasi, serta interpretasi terhadap nilai evaluasi seperti akurasi dan presisi. Pembahasan dilakukan dengan mengaitkan hasil penelitian dengan teori serta studi-studi sebelumnya untuk menjawab rumusan masalah secara komprehensif.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan, termasuk implikasi temuan terhadap bidang akademik dan praktis. Bab ini juga memuat saran yang ditujukan baik untuk pengembangan penelitian selanjutnya maupun untuk

pihak institusi pendidikan sebagai pengguna akhir dari hasil penelitian ini.

DAFTAR PUSTAKA