

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital, banyak aspek kehidupan telah berubah, termasuk pendidikan tinggi. Mengadopsi sistem informasi pembelajaran dan manajemen yang terkomputerisasi telah menghasilkan akumulasi besar data akademik. Pada umumnya, data yang mencakup rekam jejak akademik siswa, termasuk nilai mata kuliah, tingkat kehadiran, Indeks Prestasi Kumulatif (IPK), dan interaksi dalam platform pembelajaran elektronik, hanya disimpan sebagai catatan administratif atau dilaporkan secara berkala. Fenomena ini menunjukkan bahwa ada perbedaan antara mendapatkan banyak data dan mendapatkan pengetahuan strategis, yang dapat membantu pengambilan keputusan di perguruan tinggi. Padahal, pola-pola, korelasi, dan informasi penting lainnya yang dapat dioptimalkan tersembunyi di balik akumulasi data.

Teori data mining muncul sebagai hasil dari kebutuhan untuk mengubah data mentah menjadi pengetahuan yang berguna. Data mining adalah kumpulan prosedur yang digunakan untuk mengekstraksi pola-pola menarik dan informasi baru dari kumpulan data yang sangat besar (Carudin et al., 2024). Dalam proses ini, teknik dari berbagai bidang digabungkan, seperti statistika, kecerdasan buatan, dan pembelajaran mesin. Setelah penggunaan data mining di bidang pendidikan, ada bidang baru yang disebut Educational Data Mining (EDM). EDM berfokus pada metode untuk memahami perilaku belajar siswa, memprediksi prestasi akademik mereka, meningkatkan efektivitas pembelajaran, dan memberikan saran kepada manajemen sekolah (Witten et al., 2016).

Fakultas Teknik Universitas Muhammadiyah Palembang, Program Studi Teknologi Informasi, selalu berusaha untuk meningkatkan proses dan hasil pendidikan. Hasil belajar siswa, yang merupakan ukuran utama keberhasilan proses pendidikan, menjadi fokus utama. Selain pencapaian IPK, perkembangan

kompetensi kognitif, afektif, dan psikomotorik selama studi ditunjukkan sebagai hasil belajar. Hasil belajar siswa dipengaruhi oleh banyak faktor kompleks yang saling berhubungan. Faktor-faktor ini dapat dibagi menjadi kategori internal, seperti motivasi belajar, minat, dan gaya belajar; kategori eksternal, seperti metode pengajaran guru, dukungan keluarga, dan lingkungan sosial; dan kategori pendukung teknologi, seperti ketersediaan informasi dan pemanfaatan media pembelajaran digital (Bloom, 1956).

Salah satu langkah strategis yang harus diambil adalah mengidentifikasi komponen utama yang memengaruhi hasil belajar siswa dan membuat model yang dapat memprediksi potensi akademik mereka. Ini memungkinkan lembaga untuk melakukan intervensi dini yang tepat sasaran kepada siswa yang diprediksi akan menurunkan prestasi. Untuk mencapai tujuan tersebut, pendekatan Knowledge Discovery in Database (KDD) menawarkan kerangka kerja sistematis yang dimulai dengan pemilihan data, pemrosesan awal (preprocessing), transformasi, dan interpretasi pola.

Naive Bayes Classifier adalah salah satu algoritma data mining yang paling banyak digunakan untuk tugas klasifikasi dan prediksi. Teorema Bayes adalah dasar algoritma ini, yang menggunakan data yang telah diamati untuk menghitung kemungkinan suatu kejadian akan terjadi di masa depan. Metode ini memiliki banyak keunggulan, terutama karena sangat sederhana untuk dikomputerisasi, menganggap bahwa ada independensi antar fitur yang kuat (naif), dan dapat menghasilkan akurasi yang kompetitif. Keunggulan ini terutama berlaku untuk dataset dengan struktur yang sederhana dan ukuran menengah (Rish, 2001; Handayani & Pribadi, 2015). Naive Bayes telah terbukti berguna dalam akademik untuk memprediksi hasil studi, kinerja akademik, dan kelulusan siswa (Rahman & Susanto, 2020; Putri & Nugroho, 2021). Terlepas dari fakta bahwa algoritma Naive Bayes telah banyak digunakan dalam penelitian prediksi akademik, hanya ada beberapa penelitian yang secara khusus melihat bagaimana algoritma ini dapat diterapkan pada mahasiswa Fakultas Teknik Program Studi Teknologi Informasi di Universitas Muhammadiyah Palembang.

Pola dan faktor determinan yang berbeda dapat dihasilkan oleh institusi, kurikulum, dan demografi siswa.

Oleh karena itu, sebuah kajian empiris yang komprehensif diperlukan untuk menganalisis dan menerapkan metode data mining, terutama algoritma Naive Bayes, untuk memprediksi hasil belajar mahasiswa dalam konteks tertentu. Dengan menggunakan data awal yang dikumpulkan melalui kuesioner, serta data dari Sistem Informasi Akademik (SIKAD), penelitian ini bertujuan untuk mengisi kekosongan tersebut. Faktor-faktor non-akademik seperti motivasi untuk belajar, tingkat keaktifan, dan persepsi dukungan lingkungan dapat digali melalui data kuesioner. Faktor-faktor ini mungkin tidak tercatat dalam basis data formal. Diharapkan model prediksi yang lebih luas dan akurat dapat dibuat dengan menggabungkan kedua sumber data ini. Hasil analisis tidak hanya akan menghasilkan model klasifikasi dengan metrik evaluasi seperti akurasi, presisi, dan recall, tetapi juga akan mengidentifikasi faktor yang paling penting melalui analisis fitur (feature importance). Penelitian ini diharapkan dapat memberikan rekomendasi berbasis data kepada fakultas dan pengelola program studi untuk meningkatkan kualitas pembelajaran dan untuk membuat sistem pendukung keputusan akademik yang lebih strategis dan responsif.

1.2 Identifikasi masalah

Penelitian ini berangkat dari sejumlah permasalahan yang dihadapi, di antaranya:

A. Pengelolaan data akademik

Data akademik yang tersedia seringkali hanya digunakan sebagai arsip tanpa digunakan untuk analisis lebih lanjut untuk mendukung pengambilan keputusan strategis. Akibatnya, potensi data untuk memberikan wawasan berharga menjadi kurang optimal.

B. Kualitas pendidikan

Meningkatkan kualitas pendidikan, terutama dalam program studi teknologi informasi universitas muhammadiyah Palembang, membutuhkan pendekatan berbasis data.

C. Metode analisis yang efisien

Tidak ada metode analisis prediksi yang sistematis dan dapat diandalkan yang digunakan untuk menemukan pola penting dalam data akademik. Akibatnya, komponen penting yang memengaruhi hasil belajar siswa belum diungkapkan secara menyeluruh.

1.3 Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang diangkat dalam penelitian ini adalah sebagai berikut Bagaimana penerapan metode Naïve Bayes dalam analisis data akademik mahasiswa pada Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, Faktor-faktor apa saja yang signifikan dalam memprediksi hasil belajar mahasiswa, dan Sejauh mana tingkat akurasi prediksi hasil belajar mahasiswa menggunakan metode Naïve Bayes.

1.4 Tujuan penelitian

Berdasarkan rumusan masalah yang telah dibuat, maka tujuan dari penelitian ini ialah Melakukan analisis data akademik mahasiswa di Program Studi Teknologi Informasi Fakultas Teknik Universitas Muhammadiyah Palembang untuk mengidentifikasi pola-pola yang mempengaruhi hasil belajar mereka, Menggunakan metode data mining, khususnya algoritma Naive Bayes, dalam proses klasifikasi atau prediksi hasil belajar mahasiswa (seperti IPK atau status kelulusan), Membangun model prediksi hasil belajar mahasiswa yang berbasis pada algoritma Naive Bayes yang dapat membahas bagaimana hasil belajar mereka dipengaruhi oleh berbagai faktor, Mengevaluasi kinerja metode Naive Bayes dalam memprediksi hasil belajar siswa berdasarkan metrik evaluasi seperti akurasi, presisi, dan recall, dan Berdasarkan hasil analisis data, memberikan saran strategis untuk pengelola program studi atau fakultas untuk meningkatkan kualitas pembelajaran.

1.5 Kebaruan

Kebaruan yang ada pada penelitian ini meliputi beberapa aspek yang belum banyak diimplementasikan dalam penelitian sebelumnya yaitu sebagai berikut:

A. Konteks Spesifik

Penelitian ini difokuskan pada Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, yang belum banyak dilakukan kajian serupa dengan pendekatan data mining khususnya metode Naïve Bayes. Hasil penelitian akan memberikan model prediksi yang sesuai dengan karakteristik mahasiswa di lingkungan universitas ini.

B. Pemanfaatan Data Akademik Mahasiswa

Data yang digunakan mencakup berbagai variabel akademik mahasiswa (misalnya nilai mata kuliah, kehadiran, aktivitas akademik, dan faktor lainnya). Pengolahan data dilakukan melalui tahapan data preprocessing untuk menghasilkan dataset yang lebih bersih dan representatif dalam mendukung proses klasifikasi.

C. Implementasi Algoritma Naïve Bayes

Walaupun metode Naïve Bayes telah banyak digunakan dalam penelitian prediksi akademik, penerapannya dalam lingkup spesifik mahasiswa Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang masih jarang dilakukan. Penelitian ini akan menunjukkan sejauh mana akurasi, presisi, dan recall metode Naïve Bayes dalam memprediksi hasil belajar mahasiswa di konteks lokal.

D. Pengembangan Sistem Pendukung Keputusan

Hasil analisis tidak hanya berhenti pada perhitungan manual, melainkan diimplementasikan dalam bentuk sistem aplikasi sederhana yang dapat membantu pihak fakultas atau program studi dalam melakukan pemantauan dan evaluasi hasil belajar mahasiswa secara lebih efektif.

E. Manfaat Praktis

Penelitian ini memberikan kontribusi praktis berupa alat prediksi berbasis data mining yang dapat mendukung dosen dan pihak fakultas dalam melakukan intervensi dini terhadap mahasiswa yang berpotensi mengalami penurunan prestasi akademik dan Kebaruan juga terletak pada pengintegrasian analisis data

mining dengan implementasi sistem, sehingga tidak hanya menghasilkan model teoritis, tetapi juga solusi yang dapat digunakan langsung.

1.6 Manfaat penelitian

Manfaat-manfaat yang diperoleh dari hasil penelitian ini dapat dilihat dari beberapa sudut pandang, yaitu bagi penulis, pengguna sistem, dan peneliti lain.

- A. Bagi institusi pendidikan : Memberikan wawasan baru dalam pengelolaan data akademik untuk mendukung pengambilan keputusan strategis.
- B. Bagi mahasiswa : Memberikan gambaran tentang faktor-faktor yang dapat meningkatkan keberhasilan akademik.
- C. Bagi peneliti lain: Menjadi referensi bagi penelitian lebih lanjut di bidang penerapan data mining dalam pendidikan.

1.7 Pembatasan masalah

Ruang lingkup penelitian ini dibatasi pada beberapa elemen berikut agar penelitian ini dapat dilakukan dengan tepat dan mencapai tujuan yang telah ditetapkan:

- A. Mahasiswa Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang adalah subjek penelitian ini.
- B. Penelitian ini menggunakan data dari Sistem Informasi Akademik (SIKAD) dan kuesioner yang diberikan kepada siswa, yang mencakup tingkat kehadiran, nilai mata kuliah, Indeks Prestasi Kumulatif (IPK), dan jumlah SKS.
- C. Untuk memprediksi hasil belajar siswa, penelitian ini tidak membandingkan metode klasifikasi Naïve Bayes dengan algoritma klasifikasi lainnya dalam proses pemodelan.
- D. Proses penelitian terbatas pada fase data mining, yang mencakup pengumpulan data, preprocessing, pemodelan menggunakan algoritma Naive Bayes, dan evaluasi model menggunakan confusion matrix dan metrik evaluasi seperti akurasi, ketepatan, recall, dan skor F1.

Hasil penelitian fokus pada analisis prediksi hasil belajar siswa berdasarkan data akademik yang tersedia, serta mengevaluasi kinerja model dalam klasifikasi data.

1.8 Sistematika Penulisan

Penulisan Sistematis skripsi ini diartikan sebagai pedoman ataupun garis besar penulisan laporan penelitian yang dilakukan serta secara jelas mendeskripsikan isi laporan penelitian guna menjalin hubungan antara bab pertama sampai akhir.

Sistem penulisan laporan tesis ini terdiri atas:

BAB I PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan serta kaidah penelitian, metode penelitian yang digunakan, metode pengumpulan data, metode analisis data, metode pengujian serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab kedua membahas tentang landasan teori, yaitu teori-teori umum serta khusus guna membantu penulisan skripsi ini. Masalah yang meliputi di dalamnya merupakan pembahasan mengenai metode *Neve bayes*, *Algoritma C4.5*

BAB III METODOLOGI

Bab ini membahas tahapan analisis dan perancangan sistem pendukung keputusan yang digunakan untuk memprediksi hasil belajar mahasiswa. Analisis dilakukan terhadap kebutuhan fungsional dan non-fungsional sistem, serta terhadap data yang dibutuhkan dalam proses prediksi. Selanjutnya, dilakukan perancangan sistem yang mencakup perancangan alur proses, struktur data, serta penerapan algoritma *Naïve Bayes* sebagai inti dari metode klasifikasi. Perancangan ini bertujuan untuk

membangun sistem yang mampu memberikan hasil prediksi yang akurat dan dapat dimanfaatkan oleh pihak akademik dalam proses pengambilan Keputusan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan pembahasan terhadap sistem pendukung keputusan yang dibangun menggunakan algoritma *Naïve Bayes* dalam memprediksi hasil belajar mahasiswa. Pembahasan mencakup proses pengolahan data, implementasi metode, hasil prediksi, serta evaluasi kinerja sistem. Analisis dilakukan berdasarkan data mahasiswa Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, dengan tujuan untuk mengetahui sejauh mana sistem dapat memberikan prediksi yang akurat dan membantu pihak akademik dalam pengambilan keputusan terkait pembinaan dan pemantauan mahasiswa.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian yang dilakukan, serta saran-saran yang dapat dikemukakan untuk penelitian lebih lanjut atau untuk penerapan model analisis sentimen berbasis aspek.

DAFTAR PUSTAKA Bagian ini mencantumkan semua sumber referensi yang digunakan dalam penulisan penelitian, termasuk buku, artikel jurnal, dan sumber lainnya yang relevan dengan topik penelitian.