

BAB I

PENDAHULUAN

1.1 Latar Belakang

Knowledge Discovery in Database (KDD) adalah proses untuk menggali dan menganalisis sejumlah data, mengestrak informasi, dan pengetahuan yang berguna (Qisthiano et al., 2021). Teknik menganalisis data dalam penerapan *Data Mining* ini menggunakan proses tahapan *Knowledge Discovery in Databases* (KDD) yang terdiri dari *Database*, *Data Selection*, *Data Cleaning*, *Data transformation*, *Data Mining*, dan *Interpretation* (Fitrianah et al., 2021).

Data Mining adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar pengambilan keputusan. *Data mining* memproses data dalam jumlah besar menjadi informasi atau pengetahuan, diperlukan suatu teknik atau metode (Fadri, 2023). *Data Mining* memiliki beberapa teknik seperti klasifikasi, estimasi, prediksi, asosiasi, dan *clustering* (Hartati & SAN, 2022). Penelitian ini menggunakan teknik *Data Mining* klasifikasi.

Klasifikasi merupakan suatu alur proses dalam menemukan suatu fungsi atau permodelan dalam melakukan penggambaran serta menentukan kelas untuk setiap data yang bertujuan untuk memperkirakan kelas dari objek yang belum diketahui (Azhari et al., 2021). Salah satu metode yang digunakan dalam melakukan klasifikasi adalah metode *Naive Bayes* dengan menerapkan teknik pemilihan atribut atau fitur menggunakan *forward selection*.

Naive Bayes adalah suatu metode klasifikasi probabilistik yang sederhana untuk menghitung beberapa kemungkinan dengan dijumlahkannya frekuensi dan kombinasi nilai dari suatu *dataset*. Teorema Bayes diasumsikan bahwa atribut yang tidak saling bergantung (*independent*) diberikan oleh nilai pada variabel *class* (Paquin et al., 2022). Dengan menerapkan *feature forward selection* pada algoritma *Naive Bayes* yang

digunakan untuk mengetahui atribut yang paling mempengaruhi terhadap ketepatan masa studi mahasiswa (Kurniadi et al., 2022).

Berdasarkan penelitian terdahulu yang dilakukan oleh (Puspitasari et al., 2022) yang menghasilkan nilai akurasi sebesar 64,25% dan tingkat kesalahan yang rendah sebesar 35,7% menunjukkan bahwa *Naive Bayes Classifier* dapat mengklasifikasikan mutu varietas kelapa sawit dengan cukup baik. (Titimeidara & Hadikurniawati, 2021) yang menghasilkan tingkat akurasi sebesar 88 % dari 300 data yang dibagi menjadi 2 yaitu 275 data sebagai data latih dan 25 data sebagai data uji. (Hozairi et al., 2021) melakukan penelitian yang merekomendasikan bahwa model *Naive Bayes* memiliki tingkat akurasi tertinggi. (Akbar et al., 2022) yang menghasilkan tingkat akurasi tertinggi yaitu *Naive Bayes Classifier* dan *Correlation Attribute* sebesar 81.2% meningkat 12% dibandingkan tanpa *feature selection*.

Politeknik LP3I Kampus Padang adalah salah satu perguruan tinggi dengan jenjang pendidikan Diploma Tiga (D-III) Manajemen Informatika yang bernaung di bawah bendera LP3I group di Indonesia yang berpusat di Padang Sumatera Barat (Hadi, 2022). Politeknik LP3I Kampus Padang memiliki sumber daya manusia yang bekerja untuk mengoptimalkan dan meningkatkan kemajuan perguruan tinggi, sehingga mampu bersaing dengan lembaga pendidikan lainnya (Bahri, 2022). Tujuan utama lembaga pendidikan diantaranya adalah pendidikan berkualitas bagi mahasiswanya (Mildawati, 2021), oleh karena itu keterlambatan masa studi mahasiswa merupakan juga salah satu hal yang mengakibatkan penurunan kualitas pendidikan yang diberikan oleh lembaga pendidikan terhadap peserta didiknya (Izzatunnisa et al., 2021).

Mahasiswa adalah pilar penting dalam siklus hidup perguruan tinggi (Hozairi et al., 2021). Keberhasilan program studi dapat diukur dari ketepatan masa studi mahasiswa (Hasibuan & Mahdiana, 2023). Ini mengacu pada waktu yang harus ditempuh mahasiswa dari awal program studi hingga lulus, sesuai dengan jadwal yang ditetapkan oleh perguruan tinggi (Joan Galopo & Eugene S. Perez, 2021). Ketepatan masa studi sangat penting karena berdampak pada efisiensi waktu, sumber daya, dan biaya. Mahasiswa yang menyelesaikan studi tepat waktu dapat menghemat biaya pendidikan dan memasuki dunia kerja lebih cepat. Ini juga mencerminkan kemampuan mahasiswa dalam manajemen waktu dan pencapaian akademik (Kara, 2021).

Pada jenjang Diploma Tiga (D-III) dikatakan lulus tepat waktu apabila dapat menyelesaikan studi kurang atau sama dengan tiga tahun akademik dan dikategorikan lulus tidak tepat waktu apabila menyelesaikan studi lebih dari tiga tahun akademik (Kurniadi et al., 2022). Dalam perspektif yang lebih luas, ketepatan masa studi mahasiswa dapat mempengaruhi kualitas program studi, sehingga ketepatan masa studi mahasiswa dijadikan salah satu kriteria untuk menentukan penilaian akreditasi oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT). Oleh karena itu, setiap lembaga pendidikan perlu memberikan perhatian serius terhadap ketepatan masa studi mahasiswa dengan tujuan untuk membantu pihak program studi dan fakultas dalam menganalisis luaran pembelajaran dan melakukan perbaikan yang diperlukan (Hozairi et al., 2021). Penurunan tingkat kelulusan secara signifikan dapat menjadi permasalahan yang serius bagi perguruan tinggi. Hal ini tidak hanya berdampak pada prestasi akademik mahasiswa, tetapi juga dapat mempengaruhi reputasi dan akreditasi perguruan tinggi tersebut (Zulfallah, 2022). Dalam mengatasi masalah ini, prediksi ketepatan masa studi mahasiswa dapat menjadi solusi yang berguna. Dengan memahami faktor - faktor yang memengaruhi ketepatan masa studi, lembaga pendidikan dapat mengambil langkah - langkah proaktif untuk membantu mahasiswa menyelesaikan studi tepat waktu dan mengurangi tingkat kelulusan yang terlambat.

Berdasarkan uraian permasalahan diatas, maka peneliti tertarik melakukan penelitian dalam bentuk tesis dengan judul **“Implementasi Algoritma *Naive Bayes Classifier* Menggunakan *Feature Forward Selection* Dalam Memprediksi Ketepatan Masa Studi Mahasiswa”**.

1.2 Perumusan Masalah

Rumusan masalah adalah pengarah tujuan dari sebuah tulisan ilmiah agar focus terhadap pembahasan tertentu. Berdasarkan latar belakang yang telah dipaparkan, maka didapatkan rumusan masalah pada penelitian ini sebagai berikut :

1. Bagaimana penerapan algoritma *Naive Bayes* dapat memprediksi ketepatan masa studi mahasiswa ?
2. Bagaimana proses pengujian Algoritma *Naive Bayes* dengan menggunakan aplikasi *RapidMiner 9.10* dalam memprediksi ketepatan masa studi mahasiswa ?

1.3 Batasan Masalah

Batasan masalah adalah ruang lingkup masalah yang ingin dibatasi oleh peneliti yang disebabkan masalah yang terlalu luas yang bisa mengakibatkan penelitian itu tidak bisa focus. Supaya lingkup permasalahan tidak melebar, maka penulis membatasi penyelesaian permasalahan sebagai berikut :

1. Algoritma yang digunakan dalam penelitian ini adalah *Naive Bayes*
2. Data yang diolah adalah data akademik mahasiswa Politeknik LP3I Kampus Padang pada program studi Diploma Tiga (D-III) Manajemen Informatika berupa data kelulusan mahasiswa tahun 2020 - 2022.
3. Data akademik mahasiswa program studi D3 Sistem Informasi Universitas Airlangga. Jumlah data uji dalam penelitian ini adalah 159 data berupa data kelulusan mahasiswa pada tiga tahun terakhir
4. Variabel data yang di *input* ialah NIM, Nama, jenis kelamin (JK), Jurusan, Indeks Prestasi Semester (IPS), Jumlah SKS (Satuan Kredit Semester), dan Indeks Prestasi Kumulatif (IPK).
5. Dalam membuat aturan, penulis menggunakan aplikasi *RapidMiner* 9.10 untuk melakukan pengujian dalam memprediksi ketepatan masa studi mahasiswa.

1.4 Tujuan Penelitian

Tujuan dari suatu penelitian dapat untuk mengidentifikasi atau menggambarkan suatu solusi untuk suatu situasi yang mengindikasikan jenis studi yang dilakukan. Adapun tujuan yang akan diperoleh dari penelitian ini adalah sebagai berikut :

1. Memahami dan mengetahui algoritma *Naive Bayes* untuk dapat memprediksi ketepatan masa studi mahasiswa.
2. Untuk menganalisa data menggunakan algoritma *Naive Bayes* dalam memprediksi ketepatan masa studi mahasiswa.
3. Untuk merancang model yang dihasilkan dari algoritma *Naive Bayes* berdasarkan parameter yang digunakan.
4. Mengimplementasikan aplikasi *RapidMiner* untuk memprediksi ketepatan masa studi mahasiswa.

5. Untuk menguji tingkat akurasi penerapan algoritma *Naive Bayes* dalam memprediksi ketepatan masa studi mahasiswa.

1.5 Manfaat Penelitian

Manfaat penelitian adalah penegasan dan harapan peneliti, tentang hasil yang didapatkan dalam penelitian memberi manfaat atau kegunaan secara akademik dan praktis. Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

1. Diharapkan dengan adanya penelitian ini mampu untuk mengimplementasikan sistem teori tentang prediksi ketepatan masa studi mahasiswa menggunakan algoritma *Naive Bayes*.
2. Diharapkan dengan adanya penelitian ini dapat menambah referensi sebagai bahan penelitian dimasa mendatang, jika digunakan untuk memprediksi ketepatan masa studi mahasiswa menggunakan algoritma *Naive Bayes*.
3. Diharapkan dengan adanya penelitian ini dapat menambah wawasan atau informasi terkait tentang prediksi ketepatan masa studi mahasiswa menggunakan algoritma *Naive Bayes*.

1.6 Sistematika Penulisan

Sistematika penulisan merupakan sebuah metode atau urutan dalam menyelesaikan sebuah riset, penelitian, maupun karya tulis. Sistematika disesuaikan dengan template yang diatur dalam tata penulisan program studi masing – masing, seperti sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisi tentang latar belakang, perumusan masalah, batas masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini berisi tentang teori dan penerapan metode *Naive Bayes* yang digunakan dalam tahap - tahap penyelesaian masalah sesuai dengan topik penelitian.

BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan mengenai jenis penelitian yang dilakukan, pendekatan yang digunakan, sumber data, lokasi penelitian, metode penelitian, dan alat pengumpulan data, serta teknik pengolahan dan analisa.

BAB IV ANALISA DAN PERANCANGAN

Pada bab ini membahas tentang proses analisa dan perancangan sistem dengan metode *Naive Bayes* serta hasil yang didapatkan.

BAB V IMPLEMENTASI DAN HASIL

Pada bab ini membahas tentang pengujian data dan implementasi dengan menggunakan aplikasi *RapidMiner*.

BAB VI KESIMPULAN DAN SARAN

Pada bab ini membahas tentang kesimpulan dan saran dari hasil penelitian prediksi ketepatan masa studi mahasiswa yang diimplementasikan serta memberikan saran untuk peneliti selanjutnya.