

BAB I

PENDAHULUAN

1.1 Latar Belakang

Knowledge Discovery In Database (KDD) merupakan metode untuk memperoleh pengetahuan dari database yang ada. Dalam database terdapat tabel - tabel yang saling berhubungan/ berelasi. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan (*knowledge base*) untuk keperluan pengambilan keputusan (I. W. M. Putri et al., 2023). Sedangkan menurut (Ramdhani, 2023) *Knowledge Discovery in Databases* (KDD) merupakan proses mencari informasi yang lebih bernilai, lebih mudah dipahami dan baru dari penyimpanan data yang besar dan kompleks. Proses *Knowledge Discovery in Databases* (KDD) terdiri dari beberapa tahap yang sistematis, dimulai dari Pemilihan Data (*Data Selection*), Pra-pemrosesan Data (*Data Preprocessing*), Transformasi Data (*Data Transformation*), Data Mining, Evaluasi Pola (*Pattern Evaluation*), dan Presentasi Pengetahuan (*Knowledge Presentation*).

Knowledge Discovery in Database (KDD) biasanya disamakan dengan Data Mining, berdasarkan tujuannya dalam mencari pengetahuan yang berguna dalam tumpukan data yang besar. Data mining merupakan salah satu bagian proses KDD yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik (Almufqi & Voutama, 2023). Menurut

(Donga *et al.*, 2024) Data mining merupakan suatu kegiatan yang menyangkut pemakaian dan pengumpulan data untuk memperoleh hubungan atau pola dalam jumlah kumpulan data yang berukuran besar. Tujuan dari data mining adalah menemukan pola, hubungan, dan tren tersembunyi yang tidak dapat dengan mudah diidentifikasi melalui analisis manusia biasa.

Salah satu Teknik dalam data mining adalah klasifikasi. Teknik ini digunakan untuk menganalisis data yang telah dikelompokkan dan mengambil sebuah contoh (Lestari *et al.*, 2022). Sedangkan menurut (Munarsih & Ningsi, 2023) Klasifikasi adalah proses pelatihan (pembelajaran) suatu fungsi objektif (target) yang digunakan untuk memetakan setiap himpunan atribut suatu objek ke salah satu label kelas yang telah ditentukan sebelumnya. Dalam data mining terdapat banyak metode yang dapat digunakan dalam teknik klasifikasi, seperti algoritma Naïve Bayes, algoritma C4.5, dan masih banyak lainnya.

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi, segmentasi atau pengelompokan dan bersifat prediktif (Purnomo *et al.*, 2023). Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang dikembangkan oleh J. Rose Quinlan (Habibah *et al.*, 2023). Dengan adanya pengembangan tersebut, algoritma C4.5 memiliki prinsip dasar kerja yang sama dengan algoritma ID3 (Prasetyaningrum & Susanti, 2023). Prinsip kerja algoritma C4.5 adalah melakukan komputasi dengan membaca sampel pada memori dari penyimpanan dan memuatnya ke memori untuk menginstruksikan Decision Tree (Nurchahyo *et al.*, 2023).

Penelitian dengan menggunakan Algoritma C4.5 sudah banyak dilakukan oleh peneliti dengan bermacam-macam masalah. Seperti dalam penelitian yang dilakukan

oleh (Prahastiwi *et al.*, 2022), dalam Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik Menggunakan Metode Algoritma Decision Tree C4.5. Hasil akurasi yang didapat dengan menggunakan algoritma C4.5 adalah sebesar 94,11%, kemudian faktor yang menjadi root node adalah Jumlah SKS Lulus dan hasil memiliki pengaruh terhadap pemilihan peminatan.

Penelitian yang dilakukan (K. A. Putri *et al.*, 2024), dalam memprediksi masa studi mahasiswa berdasarkan data mahasiswa Universitas Muhammadiyah Prof. DR. Hamka dengan menggunakan Decision Tree Algorithm (C4.5) maka didapatkan hasil tingkat akurasi mencapai 83,87%, dengan presisi 87,50% dan recall sebesar 91,18%.

Penelitian yang dilakukan (Hasibuan & Mahdiana, 2023), dalam memprediksi ketepatan kelulusan mahasiswa berdasarkan data mahasiswa Universitas Islam Negeri Syarif Hidayatullah Jakarta dengan nilai Accuracy 75,52% akurat, Precision 75,50% aku, dan Recall 75,50%

Penelitian yang dilakukan (Sains *et al.*, 2023), dalam memprediksi predikat kelulusan mahasiswa di Politeknik Kampar dengan menggunakan algoritma C4.5. Data yang digunakan adalah data privat yang diperoleh dari Politeknik Kampar yang berisi data nilai akademik lulusan tahun 2015 – 2019 sebanyak 306 record dan data histori lulusan angkatan 2020 sebanyak 143 record. Dengan nilai akurasi sebesar 75%, rata-rata presisi sebesar 74% dan rata-rata sensitivitas sebesar 67%.

Pada Perguruan Tinggi seperti Universitas Syedza Saintika prediksi juga perlu dilakukan agar dapat menimalisir mahasiswa yang lulus tidak tepat waktu. Perguruan tinggi memiliki kewajiban untuk menghasilkan lulusan yang kompeten. Hal tersebut dapat dinilai dari tingkat kelulusan mahasiswanya. (Khasanah & Salim, 2022).

Kelulusan tepat waktu termasuk salah satu komponen pengukuran mutu lembaga pendidikan tinggi. Karena merupakan salah satu kriteria penilaian dalam proses akreditasi program studi dan institusi oleh Badan Akreditasi Perguruan Tinggi (BAN-PT) (Hasibuan & Mahdiana, 2023). Untuk alasan ini setiap lembaga pendidikan perlu memberikan perhatian serius terhadap durasi studi mahasiswa. Program studi harus secara proaktif melakukan identifikasi dini terhadap mahasiswa yang memiliki nilai akademis kurang memadai, sehingga dapat diberikan perhatian khusus dan pendampingan yang tepat guna memastikan mahasiswa tersebut dapat menyelesaikan studi sesuai waktu yang telah ditetapkan. Salah satu upaya yang dapat dilakukan adalah melalui penerapan metode prediksi kelulusan mahasiswa, yang memungkinkan pihak institusi untuk mendeteksi potensi keterlambatan studi dan melakukan intervensi yang efektif.

Dari uraian permasalahan di atas, penulis akan membahas bagaimana penerapan Algoritma Decision Tree (C4.5) dalam memecahkan masalah, dimana algoritma C4.5 digunakan untuk membangun pohon keputusan yang memetakan data ke berbagai kelas berdasarkan aturan-aturan yang ditemukan. Oleh sebab itu maka penulis tertarik melakukan penelitian dengan mengangkat judul “Implementasi Algoritma C4.5 untuk Memprediksi Tingkat Ketepatan Kelulusan Mahasiswa (Studi Kasus di Universitas Syedza Sainika)”.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah diatas, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana implementasi metode algoritma C4.5 dalam memprediksi tingkat ketepatan kelulusan mahasiswa?

2. Bagaimana menentukan faktor-faktor yang mempengaruhi tingkat ketepatan kelulusan mahasiswa?
3. Bagaimana penerapan aplikasi Rapidminer menggunakan algoritma C4.5 dalam memprediksi tingkat ketepatan kelulusan mahasiswa?

1.3 Batasan Masalah

Tujuan penulis untuk membatasi permasalahan ini agar dalam penulisan tesis ini tidak menyimpang dari rumusan permasalahan yang telah dijabarkan diatas, maka penulis merasa perlu untuk membuat suatu batasan masalah yaitu:

1. Penelitian ini difokuskan pada prediksi kelulusan mahasiswa di Universitas Syedza Saintika.
2. Penelitian ini menggunakan metode klasifikasi yaitu algoritma C4.5.
3. Faktor-faktor yang mempengaruhi tingkat kelulusan mahasiswa terbatas pada variabel yang terdapat dalam data yang tersedia.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dituangkan, maka dapat dikemukakan tujuan yang hendak dicapai pada laporan ini adalah sebagai berikut:

1. Implementasi metode algoritma C4.5 dalam memprediksi tingkat ketepatan kelulusan mahasiswa.
2. Menentukan faktor-faktor yang mempengaruhi tingkat ketepatan kelulusan mahasiswa.
3. Penerapan aplikasi Rapidminer menggunakan algoritma C4.5 dalam mengklasifikasikan tingkat ketepatan kelulusan mahasiswa.

1.5 Manfaat Penelitian

Berdasarkan rumusan masalah dan tujuan yang telah dipaparkan sebelumnya diharapkan penelitian ini memperoleh suatu manfaat. Adapun manfaat-manfaat yang diharapkan dari penelitian ini antara lain:

1. Penelitian ini dapat memberikan kontribusi pada pemahaman ilmiah tentang penggunaan metode klasifikasi dalam kelulusan mahasiswa.
2. Penelitian ini dapat membantu perguruan tinggi, khususnya Universitas Syedza Sainika dalam meningkatkan efektivitas pengambilan keputusan terkait prediksi kelulusan mahasiswa. Dengan menggunakan metode klasifikasi yang lebih akurat, perguruan tinggi dapat mengoptimalkan upaya untuk meningkatkan tingkat kelulusan mahasiswa.
3. Penelitian ini dapat membantu mengidentifikasi faktor-faktor yang mempengaruhi tingkat kelulusan mahasiswa di Universitas Syedza Sainika. Informasi ini dapat digunakan untuk mengembangkan strategi dan program pendukung yang lebih efektif guna meningkatkan tingkat kelulusan mahasiswa.

1.6 Sistematika Penulisan

Sistematika penulisan dari penelitian ini adalah sebagai berikut:

Bab I Pendahuluan

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

Bab II Landasan Teori

Bab ini membahas tentang landasan teori dan penerapan metode yang berkaitan dengan penelitian, diantaranya teori *Knowledge Discovery in Database (KDD)*, *Data*

Mining, Metode Klasifikasi, Algoritma *C4.5*, Administrasi Akademik dan Peneliti Terdahulu.

Bab III Metodologi Penelitian

Bab ini menguraikan metode yang digunakan dalam penelitian dan kerangka kerja atau tahapan-tahapan yang akan dikerjakan sampai akhir penyusunan laporan penelitian.

Bab IV Analisa Dan Perancangan

Bab ini menguraikan tentang analisa dan rancangan dalam melakukan dan menganalisa pengolahan data dan pengklasifikasian data serta bagaimana perhitungan manual algoritma *C4.5*. Pada bab ini juga akan menguraikan hasil tingkat akurasi dan *error* dari Algoritma *C4.5*.

Bab V Implementasi Dan Hasil

Bab ini membahas tentang implementasi Algoritma *C4.5* menggunakan aplikasi RapidMiner serta melihat tingkat akurasi dan *error* dari algoritma *C4.5*.

Bab VI Penutup

Bab ini berisi kesimpulan dan saran yang dihasilkan dari penelitian implementasi metode klasifikasi *Data Mining* menggunakan algoritma *C4.5* untuk prediksi ketepatan kelulusan mahasiswa.