

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kelulusan mahasiswa adalah proses yang penting dalam sistem pendidikan dan merupakan salah satu bidang yang termasuk dalam standar penjaminan mutu internal (SPMI) perguruan tinggi. Salah satu standar perguruan tinggi untuk menghasilkan mahasiswa agar lulus tepat waktu khususnya pada Program Studi S1 Teknik Informatika adalah mahasiswa dapat menempuh studi maksimal 8 semester dengan total beban studi minimal 149 sks. (Ketut Sriwinarti & luh Putu Juniarti, 2021)

Berdasarkan hasil wawancara dengan pihak Program Studi S1 Teknik Informatika Institut Teknologi PLN, penelitian ini menggunakan data mahasiswa Angkatan 2016. Pemilihan angkatan tersebut didasarkan pada ketersediaan data akademik yang paling lengkap dan jumlah data mahasiswa terbanyak dibandingkan angkatan lainnya, yaitu sebanyak 305 mahasiswa. Seluruh mahasiswa pada angkatan ini telah menyelesaikan masa studi, sehingga status kelulusan tepat waktu dan tidak tepat waktu dapat ditentukan secara final. Dari total data tersebut, sebanyak 227 mahasiswa lulus tepat waktu dan 78 mahasiswa lulus tidak tepat waktu. Kondisi ini menjadikan data Angkatan 2016 lebih representatif untuk digunakan dalam proses pemodelan dan evaluasi algoritma klasifikasi. Selain itu, penggunaan jumlah data yang lebih besar dan seimbang secara distribusi kelas diharapkan dapat meningkatkan keandalan hasil prediksi serta mengurangi potensi bias pada model. Berdasarkan informasi dari pihak Program Studi, keterlambatan kelulusan umumnya baru teridentifikasi pada semester akhir, sehingga diperlukan sistem prediksi yang mampu mendeteksi potensi keterlambatan sejak semester awal.

Salah satu teknologi yang dapat memprediksi tingkat kelulusan mahasiswa dengan menggunakan perbandingan algoritma *K-Nearest Neighbors* (KNN) dan *Decission Tree*. Tujuan membandingkan algoritma KNN (*K-Nearest Neighbors*) dan *Decision Tree* pada

prediksi kelulusan mahasiswa adalah mengetahui algoritma terbaik dengan memberikan gambaran algoritma mana yang lebih akurat, efisien, dan sesuai untuk data mahasiswa, untuk melihat algoritma model mana yang pengklasifikasiannya memiliki nilai akurasi, *precision*, dan *recall* yang lebih baik, juga diharapkan penelitian ini nantinya akan menjadi suatu tolak ukur sejauh mana tingkat kelulusan mahasiswa terutama pada mahasiswa tingkat akhir. (Putri et al., 2023)

Algoritma KNN merupakan bagian dari *Machine Learning* dimana Menggunakan komputer untuk memprediksi dan memahami sifat atau properti objek yang tidak diketahui dengan mengidentifikasi pola dalam kumpulan data. Ciri khas dari pembelajaran mesin adalah adanya proses pelatihan, pembelajaran, atau pelatihan. Dengan demikian, pembelajaran mesin memerlukan pengujian data sebagai data pelatihan, dalam hal ini yaitu Indeks Predikat Semester (IPS) 1 sampai semester 4. Penggunaan Hasil yang akan dihasilkan dari proses *training* data akan dijadikan acuan dalam menentukan apakah mahasiswa akan lulus tepat waktu atau tidak tepat waktu. (Etriyanti Sistem Informasi et al., 2021). Alasan pemilihan Algoritma KNN adalah karena memiliki kemampuan yang baik dalam menangani data numerik dan kontinu, seperti data Indeks Prestasi Semester (IPS) mahasiswa yang digunakan sebagai atribut penelitian. KNN bekerja dengan mengukur tingkat kemiripan antar data menggunakan perhitungan jarak, sehingga sangat sesuai untuk merepresentasikan pola kedekatan nilai akademik mahasiswa. KNN tidak memerlukan proses pembentukan model secara eksplisit, sehingga mampu menangkap pola data secara langsung berdasarkan distribusi data latih. Karakteristik ini menjadikan KNN efektif dalam mengidentifikasi mahasiswa yang memiliki kemiripan pola akademik, khususnya dalam mendeteksi mahasiswa yang berpotensi tidak lulus tepat waktu.

Algoritma *Decision Tree* merupakan metode pengelompokan yang paling umum digunakan, metode dalam *Machine Learning* dan data mining yang digunakan untuk membuat keputusan atau memprediksi suatu hasil dengan cara memecah data menjadi cabang-cabang berdasarkan kondisi tertentu, *Decision Tree* juga salah satu metode

pemodelan prediktif dalam analisis data yang menggunakan struktur pohon. Tujuan *Decision Tree* adalah untuk menggambarkan serta membuat keputusan berdasarkan serangkaian aturan dan kondisi. (Novianto et al., 2023). Alasan dalam pemilihan metode *Decision Tree* adalah karena memiliki keunggulan dalam menghasilkan model klasifikasi yang bersifat interpretatif dan mudah dipahami. Algoritma ini membentuk struktur pohon keputusan berdasarkan aturan-aturan yang dapat menjelaskan proses pengambilan keputusan secara transparan, sehingga cocok digunakan dalam konteks akademik dan pengambilan kebijakan. Decision Tree C4.5 mampu menangani data dengan tipe numerik maupun kategorikal melalui proses diskritisasi dan pemilihan atribut menggunakan gain ratio, sehingga dapat mengurangi bias terhadap atribut dengan banyak nilai. Keunggulan ini menjadikan C4.5 relevan untuk membangun aturan kelulusan mahasiswa yang dapat dijelaskan secara logis kepada pihak non-teknis, seperti dosen wali atau pengelola program studi.

Dalam penelitian ini, Pemilihan algoritma KNN dan Decision Tree C4.5 bertujuan untuk memperoleh perbandingan yang komprehensif antara dua pendekatan klasifikasi yang memiliki karakteristik berbeda. KNN merepresentasikan algoritma berbasis jarak (*instance-based learning*), sedangkan Decision Tree C4.5 merepresentasikan algoritma berbasis aturan (*rule-based learning*). Dengan membandingkan kedua algoritma tersebut, penelitian ini tidak hanya menilai algoritma mana yang memiliki performa terbaik dari sisi akurasi, tetapi juga mengevaluasi keseimbangan antara ketepatan prediksi dan kemudahan interpretasi model. Pendekatan ini sejalan dengan kebutuhan institusi pendidikan yang tidak hanya membutuhkan hasil prediksi yang akurat, tetapi juga dapat dipahami dan dijelaskan dalam proses pengambilan keputusan akademik.

Maka berdasarkan latar belakang di atas, penulis membuat penelitian yang berjudul **‘ANALISIS PERBANDINGAN DECISION TREE DAN K-NEAREST NEIGHBORS (KNN) DALAM PREDIKSI KELULUSAN MAHASISWA’** pada objek penelitian kelulusan mahasiswa Program Studi S1 Teknik Informatika. Hal ini

diharapkan dapat menjadi informasi yang menjadi pertimbangan serta persiapan pihak program studi untuk calon lulusan mahasiswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana kinerja algoritma *Decision Tree* dan *K-Nearest Neighbors* (KNN) dalam memprediksi kelulusan mahasiswa berdasarkan data akademik yang tersedia?
2. Metode klasifikasi mana yang memberikan hasil prediksi yang lebih akurat, antara *Decision Tree* dan *K-Nearest Neighbors* (KNN), dalam konteks prediksi kelulusan mahasiswa?
3. Apa kelebihan dan kekurangan dari masing-masing metode (*Decision Tree* dan *K-Nearest Neighbors*) dalam penerapannya terhadap data kelulusan mahasiswa?

1.3 Tujuan

Berdasarkan Rumusan Masalah yang telah diuraikan di atas, diharapkan tujuan dari penelitian ini adalah:

1. Menganalisis kinerja algoritma *Decision Tree* dan *K-Nearest Neighbors* (KNN) dalam memprediksi kelulusan mahasiswa berdasarkan data akademik yang tersedia.
2. Membandingkan tingkat akurasi dan efektivitas antara metode *Decision Tree* dan *K-Nearest Neighbors* (KNN) dalam proses prediksi kelulusan mahasiswa.
3. Mengidentifikasi kelebihan dan kekurangan dari masing-masing metode klasifikasi (*Decision Tree* dan *K-Nearest Neighbors*) dalam konteks penerapan pada data pendidikan.

1.4 Manfaat

Manfaat dari penelitian Tesis ini, yakni:

1. Memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang teknik informatika, khususnya dalam penerapan metode klasifikasi data menggunakan algoritma *Decision Tree* dan *K-Nearest Neighbors* (KNN).
2. Menjadi referensi ilmiah bagi peneliti, akademisi, dan mahasiswa yang tertarik untuk mengkaji lebih lanjut tentang penerapan algoritma data mining dalam bidang pendidikan.
3. Mendukung pengambilan keputusan berbasis data di lingkungan institusi pendidikan, khususnya dalam memprediksi dan memantau kelulusan mahasiswa secara lebih sistematis dan objektif untuk Jurusan S1 Teknik Informatika Institut Teknologi PLN
4. Mendorong pemanfaatan teknologi kecerdasan buatan dalam sistem informasi akademik, sehingga proses evaluasi dan perencanaan pendidikan dapat dilakukan secara lebih efisien dan tepat sasaran.

1.5 Ruang Lingkup Masalah

Agar penelitian ini lebih terarah dan fokus, maka ruang lingkup masalah dibatasi pada hal-hal berikut:

1. Objek penelitian terbatas pada mahasiswa Program Studi S1 Teknik Informatika Institut Teknologi PLN, yang telah menyelesaikan proses akademik.
2. Data yang digunakan adalah data primer berupa data historis akademik mahasiswa, berupa nilai Indeks Predikat Sementara (IPS) yang tersedia dalam sistem informasi akademik kampus.
3. Penelitian hanya menggunakan dua algoritma klasifikasi, yaitu *Decision Tree C4.5* dan *K-Nearest Neighbors* (KNN) dengan Nilai $K=3$, sebagai metode utama untuk membangun model prediksi kelulusan mahasiswa.

4. Evaluasi kinerja model dilakukan dengan menggunakan metrik klasifikasi, seperti akurasi, presisi, *recall*, dan *F1-score*, guna membandingkan performa kedua algoritma secara objektif.
5. Lingkup prediksi dibatasi pada status kelulusan mahasiswa, yaitu apakah mahasiswa diprediksi lulus atau tidak lulus, tanpa membahas prediksi waktu kelulusan atau faktor psikologis/non-akademik lainnya.
6. Penelitian tidak mencakup pengembangan sistem berbasis web atau integrasi langsung ke sistem informasi akademik, melainkan hanya pada pembangunan dan analisis model klasifikasi sebagai tahap awal pengembangan sistem pendukung keputusan.

1.6 Kebaruan

Penelitian ini memiliki beberapa unsur kebaruan dibandingkan dengan penelitian-penelitian sebelumnya, antara lain:

1. Konteks Institusi Khusus:

Penelitian ini dilakukan secara spesifik di Institut Teknologi PLN Program Studi S1 Teknik Informatika, yang belum banyak dijadikan objek kajian dalam penelitian serupa, sehingga memberikan kontribusi unik berdasarkan karakteristik akademik dan kurikulum lokal institusi.

2. Kombinasi dan Perbandingan Algoritma:

Penelitian ini tidak hanya menggunakan satu metode klasifikasi, tetapi secara langsung membandingkan kinerja algoritma *Decision Tree* dan *K-Nearest Neighbors* (KNN) dalam konteks prediksi kelulusan mahasiswa, sehingga memberikan perspektif yang lebih luas dan objektif terhadap efektivitas metode klasifikasi yang digunakan.

3. Program Prediksi pada Kelulusan Mahasiswa:

Fokus utama penelitian adalah prediksi kelulusan mahasiswa, bukan hanya pada performa akademik atau nilai semester, yang memberikan nilai tambah dalam

pengambilan kebijakan strategis bagi institusi pendidikan dalam rangka meningkatkan tingkat kelulusan tepat waktu.

4. Penggunaan Data Akademik Historis Nyata:

Penelitian ini menggunakan data akademik historis nyata dari mahasiswa IT PLN, bukan data simulasi, sehingga hasil analisis lebih relevan, aplikatif, dan dapat langsung digunakan oleh pihak institusi.

5. Penerapan Langsung dalam Sistem Pendukung Keputusan:

Hasil penelitian diarahkan untuk menjadi dasar pengembangan sistem pendukung keputusan (*Decision Support System*) yang dapat diintegrasikan ke dalam sistem akademik kampus guna membantu evaluasi dan pemantauan studi mahasiswa secara proaktif.

1.7 Hipotesis

Peneliti mengajukan hipotesis bahwa terdapat perbedaan kinerja antara algoritma *Decision Tree* dan *K-Nearest Neighbors* (KNN) dalam memprediksi kelulusan mahasiswa berdasarkan data akademik yang tersedia, yang ditinjau dari nilai akurasi, *presisi*, *recall*, dan *F1-score*. Selain itu, masing-masing algoritma memiliki kelebihan dan kekurangan yang berbeda dalam penerapannya pada data pendidikan, di mana algoritma KNN lebih unggul dalam ketepatan prediksi dan kemampuan mendeteksi mahasiswa yang berpotensi tidak lulus tepat waktu, sedangkan algoritma *Decision Tree* lebih unggul dalam kemudahan interpretasi model melalui aturan keputusan yang dihasilkan.

1.8 Sistematika Penulisan

Sistematika penulisan Sistem ditulis secara terstruktur dan dibagi menjadi lima bab dan dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Bab satu menjelaskan hambatan yang dapat dijelaskan dengan mendefinisikan konteks masalah dan rumusannya, menyatakan tujuan dan manfaat penelitian, ruang lingkup

masalah, kebaharuan penelitian, hipotesis dan menggambarkan sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA

Bagian ini membahas tentang teori-teori yang diambil dari literatur yang mendukung penyusunan skripsi ini. Diantara teori yang digunakan adalah teori algoritma K-Nearest Neighbor (KNN) dan *Decision Tree* C.45, lain sebagainya dan kerangka penelitian yang menggunakan diagram *fishbone* termasuk didalamnya.

BAB III METODE PENELITIAN

Bab ketiga menggunakan diagram untuk menggambarkan alur, dimulai dengan mengenali suatu masalah, mengumpulkan data, mengevaluasi kebutuhan, melaksanakan desain, proses apa yang digunakan dalam penelitian, dan kapan dan di mana penelitian dilakukan.

BAB IV HASIL DAN PEMBAHASAN

Bagian keempat merangkum temuan penelitian dan menjelaskan bagaimana hasil dari masing-masing fase penelitian diterapkan pada implementasi dan pengujian sistem.

BAB V KESIMPULAN DAN SARAN

Bab lima adalah bab terakhir di mana penjelasan yang diperoleh dari studi dan desain disajikan, bersama dengan ide-ide konstruktif untuk meningkatkan temuan penelitian masa depan.