

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG

Pendidikan tinggi, termasuk di Universitas Jambi (UNJA), memegang peranan strategis dalam pembangunan sumber daya manusia yang berkualitas. Keberhasilan mahasiswa dalam menyelesaikan studi tepat waktu merupakan salah satu indikator utama yang mencerminkan kualitas penyelenggaraan pendidikan di institusi tersebut. Berbagai faktor, seperti kemampuan akademik, latar belakang sosial-ekonomi, kehadiran, serta tingkat keterlibatan mahasiswa dalam kegiatan akademik maupun non-akademik, dapat mempengaruhi capaian kelulusan tepat waktu. Pemahaman yang komprehensif terhadap faktor-faktor tersebut memungkinkan institusi pendidikan, termasuk UNJA, untuk merumuskan strategi yang lebih efektif dalam meningkatkan tingkat kelulusan mahasiswa. Tingkat kelulusan tepat waktu juga sangat krusial bagi perguruan tinggi, karena menjadi salah satu tolok ukur dalam penilaian mutu dan peringkat institusi. Menurut Novianti & Prasetyo yang dikutip dalam penelitian Ginting dkk ada tahun 2024, Tingkat kelulusan merupakan aspek atau ukuran tingkat keberhasilan suatu perguruan tinggi dalam melaksanakan proses belajar mengajarnya [1]. Namun menurut Saputra dkk tidak semua mahasiswa dapat menyelesaikan studinya tepat waktu [2]. Oleh sebab itu, diperlukan suatu teknik untuk memprediksi waktu kelulusan mahasiswa untuk memberikan rekomendasi kepada Universitas atau Fakultas atau Program Studi dalam mengurangi mahasiswa yang lulus tidak tepat waktu.

Universitas Jambi (UNJA) merupakan salah satu perguruan tinggi negeri di Indonesia yang terdiri dari berbagai fakultas dan program studi (prodi) yang mencakup berbagai disiplin ilmu. Beberapa fakultas yang ada di UNJA meliputi Fakultas Keguruan dan Ilmu Pendidikan, Fakultas Ekonomi dan Bisnis, Fakultas Hukum, Fakultas Kedokteran dan Ilmu Kesehatan, Fakultas Pertanian, Fakultas Peternakan, serta Fakultas Sains dan Teknologi. Setiap fakultas tersebut menawarkan berbagai prodi di tingkat sarjana (S1), magister (S2), hingga doktor (S3), dengan fokus pada pengembangan pengetahuan mahasiswa di berbagai bidang untuk menghasilkan lulusan yang kompeten dan siap bersaing di dunia kerja.

Fakultas Sains dan Teknologi (SAINTEK) UNJA memiliki berbagai program studi yang berorientasi pada pengembangan ilmu pengetahuan dan teknologi. Beberapa program studi yang ditawarkan di fakultas ini mencakup Teknik Geologi, Teknik Lingkungan, Sistem Informasi, Matematika, Fisika, Kimia, dan Biologi. Setiap prodi di fakultas ini dirancang untuk membekali mahasiswa dengan pemahaman teoritis yang kuat serta keterampilan praktis yang relevan dengan kebutuhan industri dan riset. Selain itu, Fakultas Sains dan Teknologi juga aktif dalam penelitian dan pengabdian kepada masyarakat guna mendukung inovasi di bidang sains dan teknologi serta berkontribusi dalam pembangunan berkelanjutan di tingkat lokal maupun nasional. Sebagai bagian dari institusi pendidikan, Fakultas Sains dan Teknologi juga menghadapi tantangan dalam memastikan mahasiswa dapat menyelesaikan studinya tepat waktu. Faktor-faktor akademik, ketersediaan fasilitas, serta dukungan bimbingan berperan penting dalam

menentukan keberhasilan mahasiswa. Oleh karena itu, analisis terhadap lama studi mahasiswa menjadi aspek krusial dalam upaya peningkatan kualitas pendidikan.

Dalam dunia pendidikan, mahasiswa merupakan aset penting untuk dimiliki institusi penyelenggara pendidikan. Lama studi merupakan atribut penting bagi suatu universitas baik itu fakultas maupun bagi program studi. Memprediksi lama studi mahasiswa dapat dilakukan dari beberapa faktor dan kriteria yang digunakan untuk mengukur kelulusan mahasiswa [3]. Evaluasi pengukuran dilakukan guna untuk menghasilkan sarjana yang memiliki daya saing yang tinggi dan berkualitas dengan memanfaatkan basis data akademik sebagai pendukung keputusan oleh manajemen perguruan tinggi dalam meningkatkan standart kualitas. Kemajuan teknologi di bidang data mining, kecerdasan buatan, dan pembelajaran mesin membuka peluang baru dalam analisis data pendidikan. Teknologi ini memungkinkan institusi memanfaatkan tumpukan data akademik untuk mengidentifikasi pola, tren, dan faktor-faktor yang memengaruhi lama studi mahasiswa. Dengan memanfaatkan teknik data mining, institusi dapat menghasilkan informasi berharga guna menunjang kegiatan manajemen, operasional harian, dan pengambilan keputusan strategis.

Kemajuan teknologi di bidang data mining, kecerdasan buatan dan pembelajaran mesin telah membuka peluang baru dalam analisis data pendidikan terutama tingkat kelulusan. Pemanfaatan teknik data mining pada perguruan tinggi dengan memanfaatkan tumpukan data yang digunakan untuk mendapatkan informasi guna menunjang kegiatan manajemen dan kegiatan operational harian serta bermanfaat dalam mengambil keputusan [4]. Salah satu teknik dalam data

mining yaitu klasifikasi data menggunakan algoritma *Decision Tree*. Algoritma ini memiliki kemampuan untuk memproses data yang kompleks dan menghasilkan model yang dapat diinterpretasikan dengan mudah. *Decision Tree* merupakan model yang dapat digunakan dalam mengklasifikasi data yang dimana model ini menunjukkan informasi berharga mengenai faktor – faktor yang mempengaruhi kelulusan [5]. Hal ini memungkinkan penentuan pola dan tren yang relevan dengan tingkat kelulusan yang akan menghasilkan tingkat akurasi yang baik. Dalam meningkatkan tingkat akurasi dari kinerja model dapat dilakukan dengan melakukan seleksi fitur sebelum melakukan pemodelan, misalnya dengan menggunakan algoritma *chi-square* [6]. *Chi-Square* adalah salah satu algoritma yang dapat digunakan dalam meningkatkan tingkat akurasi dari kinerja permodelan dalam klasifikasi data. Uji *Chi-Square* sering diterapkan untuk mengevaluasi signifikansi suatu fitur dengan menghitung nilai statistik *Chi-Square* guna menentukan tingkat keterkaitannya dengan kelas.

Penelitian yang dilakukan oleh Qisthiano dkk pada tahun 2023, dilakukan penelitian penerapan algoritma *Decision Tree* dalam klasifikasi data prediksi kelulusan mahasiswa dengan sampel yaitu data alumni dari beberapa perguruan tinggi di Kota Palembang dengan jumlah sebesar 1739 dan dilakukan klasifikasi menggunakan tools *RapidMiner* [3]. Pada penelitian ini dilakukan klasifikasi data menggunakan algoritma *Decision Tree* memperoleh tingkat akurasi sebesar 87,93%. Pada penelitian yang dilakukan oleh Budiyantara dkk tahun 2020, dilakukan penelitian Komparasi Algoritma *Decision Tree*, Naive Bayes Dan K-Nearest Neighbor untuk memprediksi mahasiswa lulus tepat waktu menggunakan

data jurusan Teknik Informatika dan Sistem Informasi STMIK Widuri dengan menggunakan 11 atribut dalam klafikasi dengan hasil penelitian menggunakan tools *RapidMiner* diperoleh hasil akurasi dari Metode *Decision Tree* (C4.5) sebesar 98.04% pada pengujian ke 3, akurasi Metode Naïve Bayes sebesar 96.00% pada pengujian ke 4, dan akurasi Metode K-Nearest Neighbor (K-NN) sebesar 90.00% pada pengujian ke 2. tingkat performa dari suatu algoritma dapat dipengaruhi oleh dataset yang digunakan untuk pembuatan model klasifikasi [7]. Pada penelitian yang dilakukan oleh Moerdyanto dan Nuryana pada tahun 2021 berjudul Prediksi Kelulusan Tepat Waktu Menggunakan Pendekatan Pohon Keputusan Algoritma *Decision Tree* yang dipublikasikan pada *Journal of Informatics and Computer Science*. Pada penelitian ini dilakukan klasifikasi data kelulusan mahasiswa prodi Sistem Informasi, jurusan Teknik Informatika, Universitas Negeri Surabaya menggunakan algoritma *Decision Tree* yang bertujuan untuk melakukan data mining pada riwayat akademik mahasiswa dengan harapan dapat mengantisipasi dan menangani dini kemungkinan mahasiswa yang tidak lulus atau lulus tepat waktu. Dari penelitian tersebut diperoleh tingkat akurasi sebesar 75.95% [8].

Penelitian yang dilakukan oleh A.Saputra dkk pada tahun 2022 didalam penelitian Optimasi *Chi-Square* Dan Perbaikan Teknik *Prunning* Untuk Peningkatan Akurasi Algoritma C4.5 dalam model kasus prediksi keterlambatan biaya kuliah dilakukan klasifikasi data dengan tujuan untuk menentukan fitur, implementasi algoritma C4.5 dan mengevaluasi kinerja algoritma C4.5 [9]. Pada penelitian ini juga dilakukan penerapan *feature selection* menggunakan metode *Chi-Square* pada data yang berjumlah 12.408 data. Penelitian ini menghasilkan

tingkat akurasi algoritma C4.5 tanpa *feature selection* sebesar 61,40% sedangkan menggunakan *feature selection* 65,53%. Penelitian yang dilakukan oleh Darmaputra dan Nugroho pada tahun 2019 didalam penelitian yang berjudul *Student Graduation Prediction Using Algoritma K-Means With Fitur Selection Chi-Square* dilakukan penelitian menggunakan algoritma K-Means dengan *fitur selection Chi-Square* dalam memprediksi masa studi mahasiswa dengan hasil penelitian didapatkan tingkat akurasi sebesar 86,57 % dengan pengaruh *fitur selection Chi-Square* terhadap algoritma K-Means sebesar 8,85% dalam meningkatkan akurasi dari kinerja [10]. Penelitian yang dilakukan oleh Erfan dan Zaenudin pada tahun 2020 yang berjudul Penerapan Algoritma Decision Tree C4.5 Berbasis Seleksi Atribut *Chi-Squared* Untuk Klasifikasi Tingkat Pengetahuan Ibu Dalam Pemberian Asi Eksklusif Pada Bayi dilakukan penelitian untuk mengklasifikasi tingkat pengetahuan ibu dalam pemberian asi eksklusif pada bayi menggunakan algoritma Decision Tree C4.5 berbasis atribut *Chi-Square* dalam menunjang tingkat akurasi performa algoritma, dari penelitian didapatkan hasil bahwa penerapan atribut *Chi-Square* dapat meningkatkan tingkat akurasi dari algoritma dengan tingkat akurasi yang diperoleh 92,86% [6].

Berdasarkan dari permasalahan di atas serta beberapa penelitian sejenis yang dimana algoritma Decision Tree dan atribut *Chi-Square* memiliki tingkat akurasi yang baik dalam melakukan klasifikasi data, maka peneliti akan melakukan proses klasifikasi data menggunakan algoritma *Decision Tree* dan *feature selection* dengan metode *Chi-Square* dalam mengklasifikasi data tingkat kelulusan mahasiswa yang telah menempuh kuliah sarjana dengan judul penelitian “**PENINGKATAN**

ALGORITMA DECISION TREE DALAM MENGLASIFIKASI TINGKAT KELULUSAN MAHASISWA UNIVERSITAS JAMBI DENGAN SELEKSI FITUR *CHI-SQUARE*”.

1.2. PERUMUSAN MASALAH

Berdasarkan latar belakang di atas, maka dapat dirumuskan permasalahan dari penelitian ini sebagai berikut :

1. Bagaimana tingkat akurasi, presisi, *recall* dan *f1-score* Algoritma *Decision Tree* tanpa *feature selection Chi-Square* dalam mengklasifikasi tingkat kelulusan mahasiswa Universitas Jambi
2. Bagaimana tingkat akurasi, presisi, *recall* dan *f1-score* Algoritma *Decision Tree* dengan *feature selection Chi-Square* dalam mengklasifikasi tingkat kelulusan mahasiswa Universitas Jambi?

1.3. BATASAN MASALAH

Agar penelitian ini tidak terlalu luas dan menyimpang dari topik permasalahan. Maka peneliti membuat batasan masalah yang akan dibahas. Adapun batasan masalah pada penelitian ini adalah :

1. Penelitian ini menggunakan tools *Python*.
2. Data yang digunakan pada penelitian ini adalah data kelulusan mahasiswa Fakultas Sains Dan Teknologi Universitas Jambi dari tahun 2019 hingga 2024.
3. Penelitian ini menggunakan algoritma *Decision Tree* dalam mengklasifikasi tingkat kelulusan mahasiswa

4. Penelitian ini menggunakan *feature selection chi-square* dalam meningkatkan tingkat akurasi model klasifikasi
5. Pengukuran evaluasi kinerja model menggunakan *confusion matrix*.

1.4. TUJUAN PENELITIAN

Berdasarkan perumusan masalah yang telah penulis uraikan di atas, maka penelitian ini bertujuan sebagai berikut :

1. Menganalisis performa algoritma *Decision Tree* tanpa *feature selection Chi-Square* dalam mengklasifikasikan tingkat kelulusan mahasiswa Universitas Jambi berdasarkan metrik evaluasi akurasi, presisi, *recall*, dan *F1-score*.
2. Menganalisis performa algoritma *Decision Tree* dengan *feature selection Chi-Square* dalam mengklasifikasikan tingkat kelulusan mahasiswa Universitas Jambi berdasarkan metrik evaluasi akurasi, presisi, *recall*, dan *F1-score*.

1.5. MANFAAT PENELITIAN

Penelitian ini diharapkan dapat memberikan kontribusi baik dalam aspek teoritis maupun praktis sebagai berikut:

1.5.1 Manfaat Teoritis

1. Menambah wawasan dalam bidang data mining, khususnya dalam penerapan algoritma *Decision Tree* untuk klasifikasi tingkat kelulusan mahasiswa.
2. Memberikan kontribusi terhadap studi komparatif antara *Decision Tree* tanpa seleksi fitur dan *Decision Tree* dengan seleksi fitur *Chi-Square*,

sehingga dapat diketahui efektivitas metode dalam meningkatkan performa klasifikasi.

3. Menyediakan referensi bagi penelitian selanjutnya dalam pengembangan model klasifikasi mahasiswa berdasarkan berbagai faktor akademik dan non-akademik, yang dapat digunakan untuk meningkatkan akurasi.

1.5.2 Manfaat Praktis

1. Membantu institusi pendidikan, khususnya Universitas Jambi, dalam memahami pola kelulusan mahasiswa berdasarkan faktor-faktor yang berpengaruh.
2. Menyediakan informasi bagi akademisi dan pengambil kebijakan dalam merancang strategi berbasis data guna meningkatkan tingkat kelulusan mahasiswa.
3. Mempermudah proses pengambilan keputusan dalam memberikan intervensi akademik atau bimbingan kepada mahasiswa yang memiliki potensi keterlambatan dalam menyelesaikan studi, sehingga dapat meningkatkan efisiensi sistem pendidikan.

1.6. SISTEMATIKA PENULISAN

Penelitian ini terdiri dari 5 (lima) bab secara garis besar sistematika penulisan akan dijelaskan sebagai berikut :

BAB I : PENDAHULUAN

Bab ini membahas mengenai latar belakang, perumusan dan batasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan.

BAB II : LANDASAN TEORI DAN TINJAUAN PUSTAKA

Di bab ini, penulis akan membahas dasar-dasar teori dan tinjauan pustaka yang digunakan sebagai referensi dalam tesis ini, yang diambil dari berbagai buku dan jurnal. Teori-teori dasar yang akan diuraikan meliputi konsep *Data Mining*, Klasifikasi Data, algoritma *Decision Tree*, Seleksi Fitur (*Chi-Square*) dan *Python*. Bagian tinjauan pustaka mencakup penelitian-penelitian serupa yang dijadikan bahan perbandingan dengan penelitian yang dilakukan oleh penulis.

BAB III : METODE PENELITIAN

Pada bab ini penulis akan menguraikan mengenai alur penelitian, bahan yang digunakan dalam penelitian, serta metode yang akan digunakan pada saat penelitian.

BAB IV : ANALISIS DAN HASIL

Menyajikan analisis dan pemodelan metode klasifikasi *Decision Tree* dan seleksi fitur menggunakan metode *Chi-Square*, perancangan pengujian untuk menyajikan model hasil dari metode yang digunakan. Dan akan ditampilkan hasil dari analisis dan bentuk visualisasi analisis dari *tools Python* yang digunakan.

BAB V : PENUTUP

Bab ini yang berisikan kesimpulan-kesimpulan yang diambil dari hasil analisis serta saran-saran yang mencakup keseluruhan dari hasil penelitian.