

BAB I

PENDAHULUAN

A. Latar Belakang

Kemajuan teknologi informasi membuat kebutuhan akan akurasi data dalam pengambilan keputusan. Ketersediaan informasi yang tepat menjadi kunci dalam proses analisis dan evaluasi untuk menghasilkan keputusan yang optimal. Dalam prosesnya, setiap komponen data perlu dianalisis secara menyeluruh untuk menghasilkan kesimpulan yang akurat (Janu dkk., 2021).

Pengelolaan data telah berkembang di banyak industri, termasuk pendidikan. Penggunaan sistem pengelolaan data terkomputerisasi memberikan dampak positif bagi perguruan tinggi. Sistem ini membantu mengawasi berbagai aspek penting, seperti hasil pembelajaran, program akademik, dan informasi mahasiswa. Selain itu, sistem digital meningkatkan akurasi analisis dan mempercepat pengolahan data mahasiswa, serta memudahkan pemantauan dan evaluasi program pembelajaran. Dengan beralih dari manual ke digital, universitas dapat memeriksa data akademik dalam jumlah besar secara lebih efektif. (Subroto dkk., 2023). Pengelolaan data yang lebih sistematis dan terstruktur memungkinkan perguruan tinggi membuat keputusan yang lebih baik, sehingga meningkatkan kualitas pendidikan secara keseluruhan. (Armansyah & Ramli, 2022).

UIN Imam Bonjol Padang merupakan salah satu perguruan tinggi Islam di Sumatra barat. Fakultas Syari'ah (FS), yang merupakan salah satu fakultas yang telah mengelola beberapa program studi yang mengintegrasikan pada ilmu-ilmu keislaman dengan ilmu-ilmu hukum. Fakultas Syari'ah memiliki keterkaitan yang sangat erat dengan UIN karna Fakultas Syari'ah menjadi salah satu fakultas yang sangat penting dalam pengembangan Pendidikan tinggi Islam. Kurikulum yang telah dirancang mencakup berbagai aspek hukum Islam melalui program studi Hukum Keluarga, Hukum Tatanegara, Hukum Ekonomi Syari'ah, dan Perbandingan Mazhab.

Data yang telah di kumpulkan menunjukkan bahwa sebanyak 1.099 mahasiswa telah terdaftar di Fakultas Syari'ah pada angkatan 2018, 2019 dan 2020. Pemilihan data mahasiswa dari tiga angkatan yaitu 2018, 2019, dan 2020 didasarkan pada pertimbangan untuk mendapatkan data kelulusan terbaru. data dari tiga tahun terakhir ini memiliki sistem pencatatan yang lebih lengkap dan tersusun dengan baik dibandingkan data-data tahun sebelumnya. Informasi terperinci mengenai distribusi mahasiswa dapat di amati pada tabel 1.1, tabel 1.2 dan tabel 1.3 yang di sajikan di bawah ini. Pengelolaan data mahasiswa ini telah dilakukan secara sistematis melalui SIAKAD *cloud* yang telah di optimalkan oleh pihak universitas.

Tabel 1. 1 Jumlah mahasiswa fakultas syari'ah angkatan 2018

No	Program studi / Jurusan	Jumlah Mahasiswa	Lulus tepat waktu	Belum lulus
2018				
1	Hukum Keluarga	146	61	85
2	Hukum tatanegara	145	27	118
3	Hukum ekonomi syari'ah	143	70	73
4	Perbandingan mazhab	40	14	26
Total keseluruhan		477	176	301

(Sumber: (Arsip Tata Usaha UIN IB T.A 2018))

Tabel 1. 2 Jumlah mahasiswa fakultas syari'ah angkatan 2019

No	Program studi / Jurusan	Jumlah Mahasiswa	Lulus tepat waktu	Belum lulus
2019				
1	Hukum Keluarga	169	109	60
2	Hukum tatanegara	162	68	94
3	Hukum ekonomi syari'ah	199	92	107
4	Perbandingan mazhab	38	11	27
Total keseluruhan		559	280	279

(Sumber: (Arsip Tata Usaha UIN IB 2019.))

Tabel 1. 3 Jumlah mahasiswa fakultas syari'ah angkatan 2020

No	Program studi / Jurusan	Jumlah Mahasiswa	Lulus tepat waktu	Belum lulus
2020				
1	Hukum Keluarga	133	88	45
2	Hukum tatanegara	125	61	64
3	Hukum ekonomi syari'ah	121	59	62
4	Perbandingan mazhab	28	15	13
Total keseluruhan		411	224	187

(Sumber : (Arsip Tata Usaha UIN IB 2020))

Program studi Hukum Keluarga menjadi pilihan untuk melakukan prediksi kelulusan tepat waktu berdasarkan beberapa pertimbangan. Program studi ini memiliki jumlah data yang signifikan dengan total 448 mahasiswa dalam kurun waktu 3 tahun 175 mahasiswa di 2018, lalu 197 mahasiswa di 2019 dan 137 mahasiswa di 2020. Selain itu, Program studi ini menunjukkan tren peningkatan kelulusan yang signifikan, di mana pada tahun 2018 terdapat 65 dari 175 mahasiswa (37.1%) yang lulus tepat waktu, kemudian meningkat pada tahun 2019 menjadi 109 dari 197 mahasiswa (55.3%), lalu meningkat lagi pada tahun 2020 menjadi 89 dari 137 mahasiswa (64,9%) menunjukkan peningkatan persentase kelulusan sebesar 13.9%.

Data menunjukkan masih tingginya angka mahasiswa yang belum lulus tepat waktu, seperti yang terlihat pada Program studi Hukum Keluarga di mana pada angkatan 2018 terdapat 110 dari 175 mahasiswa yang belum lulus tepat waktu. Meskipun terjadi peningkatan kelulusan di angkatan 2019, masih terdapat 88 dari 197 mahasiswa yang belum lulus tepat waktu, lalu pada angkatan 2020 masih ada 48 dari 137 mahasiswa yang belum lulus tepat waktu. Data ini menunjukkan dengan masih banyak mahasiswa yang lulus tidak tepat waktu, ini mempengaruhi akreditasi program studi dan institusi, di mana tingkat kelulusan tepat waktu menjadi salah satu indikator penilaian akreditasi (Anwar dkk., 2022a). Hal ini sesuai dengan lampiran Peraturan Badan

Akreditasi Nasional Perguruan Tinggi (BAN PT) Nomor 5 Tahun 2021 Tentang Instrumen Akreditasi Program Studi *point* 56 tentang Masa Studi disebutkan bahwa poin (nilai) tertinggi yaitu 4 diperoleh apabila masa studi mahasiswa lebih dari 3,5 tahun dan kurang atau sama dengan 4,5, sehingga dapat disimpulkan bahwa kriteria kelulusan tepat waktu mengikut poin tertinggi dari penilaian BAN-PT (BAN-PT, 2021).

Ketepatan waktu kelulusan mahasiswa merupakan hal yang sulit diprediksi secara pasti. Oleh karena itu, diperlukan suatu teknik prediksi kelulusan yang dapat membantu mengidentifikasi dan mencari solusi atas permasalahan keterlambatan studi yang sering terjadi. Data akademik mahasiswa selama masa perkuliahan dapat dimanfaatkan sebagai sumber informasi yang berharga untuk dianalisis dan diolah guna memperkirakan durasi masa studi yang akan ditempuh. Hasil pengolahan data tersebut dapat menghasilkan pola dan pengetahuan yang bermanfaat untuk pengambilan keputusan di masa mendatang (Widaningsih, 2019).

Data mining hadir sebagai solusi yang efektif dalam memprediksi kelulusan tepat waktu di institusi pendidikan. Teknik ini memungkinkan instansi untuk menemukan pola tersembunyi, tren, dan korelasi yang tidak terlihat dalam kumpulan data besar dan kompleks. Penerapan data mining dalam institusi pendidikan terbukti efektif untuk meningkatkan akurasi prediksi dan menghasilkan informasi yang mendukung pengambilan keputusan strategis (Armansyah & Ramli, 2022)

Untuk mengelola data yang kompleks tersebut, data mining menggunakan berbagai teknik analisis seperti *regretion*, *asosiation*, *clustering*, dan *classification*, yang mempermudah pengolahan data agar lebih terstruktur dan mudah dipahami (Fikri Indriawan dkk., 2022). Di antara teknik-teknik tersebut, klasifikasi menjadi salah satu metode pengelolaan data yang paling sering digunakan dalam data mining karena kemampuannya dalam mengelompokkan data secara efektif. Metode klasifikasi bertujuan untuk membangun model atau fungsi yang dapat memetakan input data ke dalam

salah satu dari beberapa kelas yang telah ditentukan. Dalam penerapannya pada *machine learning* untuk tujuan prediksi, terdapat beberapa metode klasifikasi yang umum digunakan, seperti *Decision Tree (C4.5)*, *Multilayer Perceptron Backpropagation*, *K-Nearest Neighbor*, *Naïve Bayes*, dan berbagai metode klasifikasi lainnya (Armansyah & Ramli, 2022). Algoritma canggih dan teknik pembelajaran mesin yang digunakan dalam data mining menjadikannya alat yang kuat untuk berbagai bidang, seperti bisnis, kesehatan, pendidikan, dan pemerintahan (Supardi dkk., 2023).

C4.5 merupakan salah satu metode pembelajaran mesin yang efektif untuk membangun model keputusan berbasis pohon (Suriani Uci, 2023). *Decision Tree* atau pohon keputusan menjadi pilihan populer karena kemudahannya dalam interpretasi data. Selain C4.5, terdapat beberapa algoritma pembentukan pohon keputusan seperti ID3 dan *CART*. Keunggulan algoritma C4.5 terletak pada kemampuannya dalam menangani data numerik dan kategorial, serta mengatasi masalah *missing Values* (Adha dkk., 2023).

Salah satu algoritma klasifikasi lainnya adalah *Naïve Bayes*. *Naïve Bayes* adalah salah satu metode klasifikasi dalam *machine learning* yang menggunakan pendekatan *probabilistik* berdasarkan *Teorema Bayes* (Wati & Rudianto, 2022). Metode *Naïve Bayes* merupakan metode yang memerlukan jumlah data *training* dan data *testing* yang kecil untuk memilih parameter yang dibutuhkan selama proses pengklasifikasian. Dengan menggunakan beberapa atribut yang saling berhubungan untuk penentuan kelayakan (Pebdika dkk., 2023).

Sebelumnya banyak penelitian yang menggunakan Algoritma C4.5 dan *Naïve Bayes* untuk memprediksi kelulusan tepat waktu. Seperti penelitian (Tyas dkk., 2021) melakukan analisis perbandingan algoritma *Naïve Bayes* dan C.45 dalam klasifikasi data mining untuk memprediksi kelulusan menunjukkan keefektifan data set yang sedang diolah yang diterapkan mencapai 94%. Sedangkan pada algoritma C.45 mendapatkan hasil pengukuran akurasi dalam memprediksi kelulusan tepat waktu yaitu sebesar 92,60%. Maka hal ini

menunjukkan bahwa untuk memprediksi kelulusan algoritma *Naïve Bayes* memiliki klasifikasi tingkat keakuratan yang lebih tinggi dibandingkan algoritma C.45.

Adapun penelitian lainnya yang dilakukan oleh (suriani uci, 2023) menunjukkan efektivitas tinggi algoritma C4.5 dalam memprediksi kelulusan mahasiswa. Menggunakan atribut IPK dan IPS semester 1,2,3,4 sampai dengan 7 dari 1686 mahasiswa, model mencapai akurasi 99.64% dengan nilai AUC 99.5%. Secara keseluruhan, algoritma C4.5 terbukti sangat efektif untuk memprediksi kelulusan mahasiswa berdasarkan data akademik. Dengan pendekatan analisis yang akurat dan kemampuannya dalam mengolah data kompleks serta menghasilkan model prediksi yang mudah dimengerti dan ditafsirkan, algoritma C4.5 menawarkan solusi yang efektif untuk tantangan prediksi kelulusan tepat waktu mahasiswa Program Studi Hukum Keluarga.

Dalam konteks prediksi kelulusan mahasiswa, implementasi data mining telah menunjukkan hasil yang menjanjikan. Metode *Naïve Bayes*, misalnya pada penelitian (Armansyah & Ramli, 2022), menunjukkan performa metode *Naïve Bayes* bekerja sangat baik bahkan menunjukkan akurasi hingga 100%. Perolehan itu dikarenakan konfigurasi atribut data latih yang sesuai pada rancangan model.

Kedua metode dipilih karena kelebihanannya yang mudah dipahami, memiliki perhitungan sederhana, dan cepat dalam komputasi. Pemilihan algoritma ini didukung dengan penelitian (Tyas dkk., 2021) yang mengatakan bahwa algoritma *Naïve Bayes* dan C.45 saat ini memang sedang populer hal ini karena tingkat akurasi dari kedua algoritma ini tinggi.

Penelitian ini menggunakan atribut IPK, SKS, dan IPS. Berdasarkan hasil wawancara dengan mahasiswa dan bagian akademik, IPK menjadi salah satu faktor utama yang mempengaruhi keterlambatan kelulusan mahasiswa. Pemilihan atribut ini diperkuat oleh penelitian Aji dkk (2023) yang menggunakan variabel jenis kelamin, status mahasiswa, status menikah, umur,

Indeks Prestasi Semester (IPS) dari semester 1 hingga 8, dan Indeks Prestasi Kumulatif (IPK).

Adapun penelitian yang dilakukan Anwar dkk (2022) menggunakan atribut berupa, Jenis Kelamin, Asal SLTA, Status Domisili, Penghasilan Orang Tua / Wali, Indeks Prestasi, Jumlah SKS. Dan dalam penelitian (Kamil dkk., 2020) melakukan prediksi kelulusan tepat waktu menggunakan atribut seperti IPK, kehadiran perkuliahan, partisipasi dalam praktikum, kegiatan penelitian, serta latar belakang pendidikan dan ekonomi mahasiswa menjadi variabel penting yang perlu dianalisis. Dalam penelitian (Armansyah & Ramli, 2022) mengatakan faktor yang memengaruhi kelulusan mahasiswa dalam penelitian ini dipicu oleh atribut indeks prestasi (IP) semester 1, 2, 3, dan 4 dengan kriteria indeks prestasi $\geq 3,00$, jumlah SKS Semester 4 dengan kriteria ≥ 80 sks, lama masa TA ≤ 1 semester, dan lama masa STUDI ≤ 4 tahun.

Berdasarkan latar belakang tersebut, peneliti tertarik untuk melakukan penelitian dengan judul **Perbandingan Metode C4.5 Dan Naïve Bayes Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu Pada Program studi Hukum Keluarga UIN Imam Bonjol Padang**

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah

- 1) Bagaimana implementasi algoritma C4.5 dan *Naïve Bayes* dalam prediksi kelulusan mahasiswa tepat waktu pada Program studi Hukum Keluarga?
- 2) Bagaimana hasil evaluasi performa perbandingan algoritma C4.5 dan *Naïve Bayes* dalam prediksi kelulusan mahasiswa tepat waktu pada program studi Hukum Keluarga?

C. Batasan Masalah

Untuk memfokuskan penelitian, beberapa batasan masalah ditetapkan sebagai berikut:

1. Algoritma yang di bandingkan adalah C4.5 dan *Naïve Bayes* sebagai metode analisis.
2. Atribut yang digunakan meliputi: Satuan Kredit Semester (SKS), Indeks Prestasi Kumulatif (IPK), dan Indeks Prestasi Semester (IPS)
3. Data untuk analisisnya diperoleh dari Program Studi Hukum Keluarga, UIN Imam Bonjol Padang.
4. Data set yang digunakan adalah data mahasiswa program studi Hukum Keluarga dengan angkatan 2018, 2019, dan 2020.
5. Bahasa pemrograman *Python* digunakan untuk pengolahan dan visualisasi data.

D. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengimplementasikan algoritma C4.5 dan *Naïve Bayes* untuk memprediksi kelulusan mahasiswa tepat waktu pada Program Studi Hukum Keluarga.
2. Mengevaluasi dan membandingkan performa algoritma C4.5 dan *Naïve Bayes* dalam memprediksi kelulusan mahasiswa tepat waktu pada Program Studi Hukum Keluarga.

E. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Teoritis

Memberikan pengetahuan baru dan pemahaman yang lebih kepada penulis tentang prediksi kelulusan mahasiswa menggunakan metode C4.5, serta di harapkan menjadi referensi untuk penelitian-penelitian selanjutnya.

2. Manfaat praktis

- a. Bagi penulis

1. Penelitian ini dapat membantu penulis dalam pengembangan keilmuan tentang data mining mengacu pada data-data yang selama.
 2. Penelitian ini sebagai acuan bagi penulis untuk penelitian berikutnya yang berkenaan dengan data mining dan metode C4.5.
 3. Penelitian ini dapat menambah pengetahuan penulis dalam memecahkan apakah metode C4.5 dapat di terapkan pada permasalahan-permasalahan prediksi kelulusan mahasiswa tepat waktu melalui data yang tersedia.
 4. Penelitian ini membantu mengetahui prosedur analisis dengan metode C4.5 dalam melakukan prediksi kelulusan mahasiswa.
- b. Bagi program studi
1. Penelitian ini dapat memberikan kontribusi kepada program studi Hukum Keluarga untuk membantu pihak akademik dalam membuat kebijakan dalam meningkatkan jumlah kelulusan tepat waktu pada Program Studi Hukum Keluarga UIN Imam Bonjol Padang.
 2. Penelitian ini mampu membantu pihak akademik dalam mencari solusi dari permasalahan keterlambatan kelulusan mahasiswa Program Studi Hukum Keluarga UIN Imam Bonjol Padang.
- c. Bagi perguruan tinggi
- Penelitian ini di harapkan dapat menjadi arsip dokumen akademik Universitas Islam Negeri Imam Bonjol Padang dan juga dapat di jadikan sebagai acuan bagi mahasiswa yang ingin melanjutkan penelitian khususnya pada prediksi kelulusan mahasiswa tepat waktu pada Program studi Hukum Keluarga UIN Imam Bonjol Padang menggunakan metode C4.5.