

BAB I

PENDAHULUAN

1.1 Latar Belakang

Ketepatan waktu lulus mahasiswa menjadi salah satu indikator penilaian akreditasi program studi sebagai unit pelaksana pendidikan pada perguruan tinggi. Hal tersebut sesuai dengan peraturan Badan Akreditasi Nasional Perguruan Tinggi Nomor 3 tahun 2019 tentang Instrumen Akreditasi Perguruan Tinggi yang menyatakan bahwa salah satu indikator penilaian akreditasi adalah persentase lulusan tepat waktu untuk setiap program studi dari perguruan tinggi (BAN-PT, 2020). Pemerintah melalui DIKTI telah mengeluarkan aturan batas maksimal lama kuliah yang bisa ditempuh oleh seluruh tingkatan sarjana. Rentang waktu studi mahasiswa program strata-1 (S1) di perguruan tinggi adalah dari empat sampai tujuh tahun untuk menyelesaikan beban belajar 144 SKS. Mahasiswa lulus tepat waktu adalah mahasiswa yang mampu menyelesaikan studinya di perguruan tinggi selama kurang dari atau sama dengan empat tahun. Sedangkan, mahasiswa tidak lulus tepat waktu adalah mahasiswa menyelesaikan studinya lebih dari empat tahun.

Menurut data Biro Akademik, Kemahasiswaan, Perencanaan dan Kerjasama, Universitas Siliwangi, data ketepatan waktu lulus mahasiswa tepat waktu di Universitas Siliwangi adalah sekitar 35% untuk angkatan 2018, 36% untuk angkatan 2019 dan 38% untuk angkatan 2020 (Universitas Siliwangi, 2022). Data tersebut menunjukkan bahwa pencapaian ketepatan lulus tepat waktu mahasiswa Universitas Siliwangi belum memenuhi target pada pencapaian visi dan misi Universitas Siliwangi sebesar $\geq 55\%$ (Universitas Siliwangi, 2022) dan standar penilaian akreditasi unggul dari BAN PT sebesar $\geq 50\%$ (BADAN AKREDITASI NASIONAL PERGURUAN TINGGI, 2021). Efek jangka panjang dari ketidaktercapaian standar tersebut dapat mengakibatkan berkurangnya reputasi akademik universitas dan mengurangi daya tampung mahasiswa baru. Dari sudut pandang mahasiswa, keterlambatan

kelulusan ini berdampak pada sisi finansial dimana secara ekonomi akan menambah pengeluaran untuk membayar UKT (Uang Kuliah Tunggal) pada semester tambahan dan menyebabkan terlambatnya memasuki dunia kerja. Selain itu, bagi sebagian mahasiswa yang kelulusannya terlambat akan memberi efek secara psikologis ketika berhadapan dengan orang tua atau wali yang membiayainya, dan berinteraksi dengan keluarga, teman, tetangga dan orang lainnya. Bagi perguruan tinggi, jika jumlah mahasiswa yang lulus tepat waktu bisa dijaga mendekati 100% maka akan tercipta keseimbangan antara jumlah mahasiswa masuk dengan jumlah mahasiswa yang lulus. Keseimbangan ini penting untuk dijaga dalam pengelolaan program studi di suatu perguruan tinggi. Berbagai masalah yang mungkin dihadapi antara lain penambahan jadwal mata kuliah tertentu bagi yang mengulang dan dosen membimbing tugas akhir melebihi jumlah idealnya yang diatur dalam penilaian akreditasi program studi yang pada akhirnya akan mempengaruhi penilaian aspek-aspek yang dalam akreditasi program studi seperti rasio dosen dengan mahasiswa yang tidak ideal. Oleh karena itu, perlu adanya solusi terhadap masalah waktu kelulusan mahasiswa yang terlambat sehingga menjadi bahan evaluasi bagi Universitas Siliwangi untuk mengambil langkah proaktif dalam pencegahan dan penanganan sedini mungkin. Evaluasi ini diperlukan untuk mengurangi tingkat keterlambatan kelulusan mahasiswa, dan nantinya dapat menjadi acuan dalam pengukuran kualitas mahasiswa agar tercipta keseimbangan.

Salah satu cara yang dapat dilakukan untuk mengatasi masalah kelulusan tepat waktu yang tidak tercapai adalah dengan memanfaatkan data akademik untuk memprediksi apakah mahasiswa yang bersangkutan akan lulus tepat waktu atau tidak. Beberapa teknik untuk memprediksi kelulusan mahasiswa tepat waktu adalah menggunakan metode menggunakan model *machine learning*. Beberapa metode seperti model *K-Nearest Neighbor (KNN)* (Priyambodo et al., 2022), *C4.5* (Astuti, 2017) dan *support vector machine* (Saire-Peralta & Córdova-Martínez, 2023)(Imam Riadi et al., 2024), dan *random forest* (Mu'tashim & Zaidiah,

2023)(Saire-Peralta & Córdova-Martínez, 2023) telah diterapkan untuk memprediksi kelulusan kuliah tepat waktu. Hasil penelitian menunjukkan bahwa ketika model-model tersebut dihadapkan pada persoalan dataset dengan fitur yang lebih besar dan tidak seimbang menjadi bermasalah dengan hanya mendapatkan akurasi yang rendah. Disamping itu, penelitian-penelitian tersebut lebih banyak menganalisa data numerik tanpa menyeimbangkannya dengan data kategorikal, sehingga mengganggu prediksi.

Model XGBoost adalah algoritma yang berbasis pada pohon keputusan yang mempunyai kemampuan menangani data yang besar, tidak seimbang dan kompleks serta menyediakan parameter regularisasi untuk mencegah *overfitting*. Implementasi metode XGBoost telah diterapkan pada 291 data dengan enam atribut yaitu asal SLTA, program studi, provinsi, rata-rata matematika, IPK dan lama studi dan itu menemukan akurasi perhitungan sebesar 72% (Imam Riadi et al., 2024). Namun, model ini masih mempunyai kesalahan klasifikasi cukup besar yang terjadi karena atribut yang digunakan kurang bervariasi dalam pencapaian indek prestasi tiap semester. Selanjutnya, perhitungan menggunakan metode XGBoost ini diterapkan untuk 5.408 data dengan fitur data performa akademik mahasiswa selama empat semester dan menemukan akurasi perhitungan yang hampir sama sebesar 74,32% (Heriqbaldi, 2024). Pada penelitian lain, metode ini digunakan untuk perhitungan pada 649 data dengan atribut input data akademik enam semester dan penghasilan orang tua dan menemukan perubahan positif dalam akurasi sebesar 82,9% (Muhammady et al., 2024). Selanjutnya, penerapan model XGBoost untuk memprediksi kelulusan tepat waktu mahasiswa telah dilakukan pada dataset 13.377 mahasiswa yang menunjukkan penggunaan 12 atribut input performa akademik yang terdiri dari jumlah SKS 8 semester dan IPK 8 semester memiliki akurasi tinggi sebesar 92,18% (Nirmala et al., 2022). Model ini terbukti memiliki akurasi dan kecepatan tinggi saat diaplikasikan ke dalam basis data dengan dataset besar dan atribut input besar. Berdasarkan hasil investigasi dalam beberapa penelitian yang menerapkan model

XGBoost, hasil perangkingan atribut menunjukkan bahwa atribut input memiliki pengaruh besar dalam klasifikasi kelulusan kuliah tepat waktu mahasiswa. Hasil ini sejalan dengan hasil evaluasi hasil kinerja algoritma XGBoost yang diterapkan pada prediksi banjir dan status stanting berturut-turut menghasilkan nilai akurasi sebesar 98,86% dan 96,5% (Asnawi et al., 2024)(Muhamad Fikri, 2024).

Dari penelitian yang telah dilakukan, terbukti bahwa metode XGBoost mampu meningkatkan akurasi dalam membangun model prediksi, terutama ketika diterapkan pada dataset berukuran besar dengan banyak fitur. Namun demikian, dalam beberapa studi terdahulu ditemukan bahwa peningkatan jumlah fitur tidak selalu berbanding lurus dengan kualitas prediksi. Justru, terdapat kemungkinan bahwa fitur-fitur yang kurang relevan dapat menyebabkan noise, memperlambat proses pelatihan, dan bahkan menurunkan performa model. Oleh karena itu, diperlukan eksplorasi mendalam terhadap pentingnya setiap fitur, salah satunya melalui teknik feature importance. Analisis feature importance pada XGBoost memberikan gambaran global mengenai kontribusi relatif masing-masing fitur terhadap keseluruhan prediksi model. Dengan teknik ini, peneliti dapat melakukan seleksi fitur yang lebih tepat sasaran guna menyederhanakan model tanpa kehilangan akurasi yang signifikan.

Lebih lanjut, untuk memberikan pemahaman yang lebih detail terhadap kontribusi setiap fitur pada prediksi individu, dapat digunakan metode SHAP (SHapley Additive exPlanations)(I.U. Ekanayake, D.P.P. Meddage, n.d.). SHAP bekerja dengan menghitung dampak setiap fitur terhadap perubahan output model, sehingga memberikan interpretasi yang lebih transparan dan dapat dipercaya. Analisis SHAP tidak hanya menunjukkan fitur mana yang paling penting secara keseluruhan, tetapi juga bagaimana arah pengaruh fitur tersebut—apakah meningkatkan atau menurunkan probabilitas suatu prediksi, baik untuk setiap mahasiswa secara individual maupun secara agregat. Dengan kombinasi kedua metode ini, proses interpretasi hasil model menjadi lebih komprehensif dan tidak hanya berfokus pada kinerja

angka semata, tetapi juga pemahaman mendalam terhadap faktor-faktor yang memengaruhi prediksi.

Melalui analisis feature importance dan SHAP, model XGBoost tidak hanya menjadi alat prediksi yang kuat, tetapi juga dapat berfungsi sebagai alat diagnosis yang membantu institusi pendidikan mengidentifikasi fitur-fitur utama yang berpengaruh terhadap kelulusan mahasiswa. Hal ini memungkinkan pengambilan kebijakan yang lebih tepat sasaran, seperti peningkatan bantuan kepada mahasiswa dengan kondisi sosial-ekonomi tertentu atau evaluasi efektivitas jalur masuk. Eksplorasi lanjutan terhadap fitur yang relevan diharapkan dapat mengoptimalkan kinerja model, baik dari segi kecepatan pelatihan maupun efektivitas hasil prediksi.

1.2 Rumusan Masalah

1. Bagaimana model XGBoost dapat memprediksi ketepatan waktu kelulusan mahasiswa?
2. Apa saja faktor-faktor yang mempengaruhi ketepatan waktu kelulusan mahasiswa?

1.3 Tujuan Penelitian

1. Mengembangkan model prediksi menggunakan algoritma XGBoost untuk menilai ketepatan waktu kelulusan mahasiswa.
2. Mengidentifikasi dan menganalisis faktor-faktor yang berpengaruh terhadap ketepatan waktu kelulusan mahasiswa.

1.4 Manfaat Penelitian

1. Memberikan wawasan yang lebih dalam tentang faktor-faktor yang mempengaruhi ketepatan waktu kelulusan mahasiswa, yang dapat digunakan oleh institusi pendidikan untuk meningkatkan strategi pembelajaran.
2. Menyediakan model prediksi yang dapat diimplementasikan oleh perguruan tinggi untuk memantau dan mendukung mahasiswa yang berisiko tidak lulus tepat waktu.

3. Berkontribusi pada pengembangan literatur mengenai penerapan machine learning dalam pendidikan, khususnya dalam konteks prediksi kelulusan.

1.5 Batasan Masalah

1. Penelitian ini hanya akan fokus pada mahasiswa di satu atau beberapa perguruan tinggi tertentu, sehingga hasilnya mungkin tidak dapat digeneralisasi untuk semua institusi pendidikan.
2. Data yang digunakan dalam penelitian ini terbatas pada variabel-variabel yang tersedia dalam sistem informasi akademik, sehingga faktor-faktor lain yang mungkin berpengaruh tidak akan dianalisis.
3. Model yang dikembangkan dalam penelitian ini hanya akan menggunakan algoritma XGBoost, tanpa membandingkannya secara mendalam dengan algoritma machine learning lainnya, meskipun analisis akurasi akan dilakukan.
4. Analisa terhadap model yang dikembangkan akan berfokus pada hal-hal substansial yang berkaitan antara data dengan hasil prediksi model.