

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Kelulusan mahasiswa tepat waktu merupakan salah satu indikator penting dalam menilai kualitas pendidikan tinggi, karena hal ini mencerminkan kemampuan institusi pendidikan dalam memberikan pendidikan yang efektif dan efisien. Universitas Islam Riau (UIR), sebagai salah satu institusi pendidikan tinggi di Indonesia, memiliki tujuan yang sama, yaitu meningkatkan kualitas pendidikan dan meningkatkan tingkat kelulusan mahasiswa tepat waktu. Namun, dalam kenyataannya, masih banyak mahasiswa yang tidak lulus tepat waktu, yang tidak hanya mempengaruhi reputasi universitas, tetapi juga meningkatkan biaya pendidikan dan menghambat pencapaian tujuan pendidikan yang lebih luas.

Ketepatan waktu kelulusan merupakan salah satu indikator penting dalam mengevaluasi efektivitas institusi pendidikan tinggi [1]. Universitas diharapkan dapat memastikan bahwa sebagian besar mahasiswa menyelesaikan studinya dalam waktu yang telah ditentukan [2]. Keterlambatan kelulusan dapat menghambat proses akreditasi institusi, mengganggu perencanaan akademik, dan mengurangi ketersediaan lulusan yang tepat waktu untuk memenuhi kebutuhan pasar kerja [3]. Oleh karena itu, deteksi dini terhadap potensi keterlambatan kelulusan sangat penting untuk implementasi intervensi proaktif.

Masalah ketidakseimbangan kelas (class imbalance) merupakan salah satu tantangan dalam pembelajaran mesin (machine learning) untuk memprediksi kelulusan mahasiswa tepat waktu. Ketidakseimbangan kelas terjadi ketika jumlah data yang tersedia untuk kelas minoritas (mahasiswa yang tidak lulus tepat waktu) jauh lebih sedikit dibandingkan dengan kelas mayoritas (mahasiswa yang lulus tepat waktu). Hal ini dapat menyebabkan model pembelajaran mesin menjadi bias terhadap

kelas mayoritas dan memiliki kinerja yang kurang optimal dalam memprediksi kelas minoritas.

Kemajuan teknologi telah memungkinkan integrasi metode pembelajaran mesin dalam pendidikan, termasuk prediksi hasil kelulusan mahasiswa [4]. Algoritma seperti Decision Tree, Naive Bayes, K-Nearest Neighbors (K-NN), dan Support Vector Machine (SVM) telah banyak digunakan karena kemampuannya mengidentifikasi pola dalam data historis mahasiswa [5], [6], [7], [8]. Namun, banyak dataset pendidikan yang mengalami ketidakseimbangan kelas, di mana jumlah lulusan tepat waktu jauh lebih banyak daripada mereka yang terlambat lulus [9]. Ketidakseimbangan ini dapat mempengaruhi kinerja klasifikasi, sehingga prediksi cenderung bias terhadap kelas mayoritas dan mengabaikan mereka yang berisiko tidak lulus tepat waktu [10].

Oleh karena itu, perlu dilakukan penelitian untuk mengatasi masalah ketidakseimbangan kelas dalam pembelajaran mesin guna memprediksi kelulusan mahasiswa tepat waktu di UIR. Penelitian ini bertujuan mengembangkan model pembelajaran mesin yang dapat memprediksi kelulusan mahasiswa tepat waktu dengan akurat, sehingga universitas dapat mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu dan memberikan intervensi yang tepat untuk meningkatkan kemungkinan lulus tepat waktu.

Beberapa penelitian sebelumnya telah menerapkan pembelajaran mesin untuk memprediksi hasil kelulusan. Misalnya, penelitian [11] menggunakan Decision Tree dan Naïve Bayes berdasarkan kinerja akademik, namun tidak mempertimbangkan distribusi kelas. Penelitian [12] mengeksplorasi berbagai algoritma, termasuk DT, SVM, Random Forest, ANN, KNN, dan Logistic Regression, namun kinerja untuk kelas minoritas masih kurang optimal. Penelitian [13] juga melaporkan kesalahan klasifikasi kelas minoritas meskipun menggunakan Random Forest. Sementara itu, penelitian [14] menunjukkan bahwa penerapan SMOTE meningkatkan akurasi klasifikasi dengan mengatasi ketidakseimbangan data secara efektif.

Tidak seperti penelitian sebelumnya, penelitian ini berfokus pada evaluasi pengaruh ketidakseimbangan kelas terhadap kinerja beberapa algoritma pembelajaran mesin menggunakan dataset spesifik Universitas Islam Riau. Yang membedakan penelitian ini adalah penekanan pada kinerja klasifikasi yang adil melalui penerapan SMOTE dan investigasi algoritma mana yang paling mendapat manfaat dari penyeimbangan tersebut. Selain itu, penelitian ini menggunakan dataset nyata yang dikumpulkan dari 120 responden (mahasiswa dan alumni) melalui kuesioner terstruktur, menyediakan informasi kaya konteks dan autentik untuk pemodelan prediktif.

Penelitian ini juga bertujuan untuk melangkah lebih jauh dari klasifikasi sederhana dengan menganalisis apakah fitur-fitur tertentu—seperti IPK, dukungan keluarga, atau kebiasaan belajar—berkorelasi dengan keterlambatan kelulusan. Wawasan yang lebih dalam ini menawarkan kontribusi baru tidak hanya dalam kinerja prediksi, tetapi juga dalam mengarahkan intervensi strategis. Dengan menerapkan tahapan pra-pemrosesan seperti penanganan nilai yang hilang, normalisasi, pengkodean, dan penyeimbangan berbasis SMOTE, data dimodelkan menggunakan empat klasifikasi: Decision Tree, Gaussian Naive Bayes, K-Nearest Neighbors, dan Support Vector Machine. Setiap algoritma dipilih karena kekuatannya yang unik: interpretabilitas [15], kesederhanaan [16], pembelajaran berbasis instansi [17], dan ketahanan dalam ruang dimensi tinggi [18].

Hasil penelitian ini diharapkan dapat memberikan informasi untuk pengembangan alat prediksi kelulusan yang adil dan dapat diandalkan, yang dapat membantu universitas dalam melakukan intervensi dini bagi mahasiswa yang berisiko tidak lulus tepat waktu, sehingga mendukung upaya institusi untuk meningkatkan tingkat keberhasilan mahasiswa.

Penelitian sebelumnya telah menunjukkan berbagai pendekatan untuk memprediksi kelulusan tepat waktu mahasiswa. Beberapa penelitian menggunakan algoritma Support Vector Machine (SVM), mencapai akurasi sekitar 71,20%, namun menghadapi tantangan dengan data yang tidak seimbang, terutama dalam

memprediksi mahasiswa yang tidak lulus tepat waktu. Sementara itu, penelitian lain menggunakan Random Forest dan Neural Networks pada dataset besar (132.734 mahasiswa), mencapai akurasi 76% dan AUC 79%, dengan IPK diidentifikasi sebagai faktor paling berpengaruh [19]. Penelitian tentang loyalitas mahasiswa menggunakan Random Forest juga menunjukkan hasil yang kuat (akurasi 90,9%) berdasarkan data kuesioner terkait persepsi kualitas layanan kampus [20].

Pendekatan lain, seperti Decision Tree dan Naive Bayes, juga memberikan hasil yang kompetitif. Sebuah penelitian menggunakan Decision Tree untuk memprediksi kelulusan di Universitas Dian Nuswantoro mencapai akurasi tinggi, yaitu 91% [21]. Naive Bayes, menggunakan 14 variabel akademik dan demografis, mencatat akurasi 85% dengan f1-score tinggi untuk prediksi kelulusan tepat waktu, namun kinerja yang lebih rendah dalam mengidentifikasi keterlambatan [22]. Beberapa penelitian menggabungkan AdaBoost dengan Decision Tree, mencapai f1-score hingga 0,82 [4]. Bahkan dalam skenario data yang tidak seimbang, Decision Tree yang ditingkatkan dengan SMOTE menunjukkan kinerja yang sangat baik (akurasi 96,67%) [23]. Selain itu, penelitian menggunakan pendekatan penambahan data pendidikan dengan Decision Tree untuk mata kuliah inti komputasi mencapai akurasi 88,9%. Data historis mahasiswa dari beberapa angkatan juga telah berhasil dianalisis menggunakan berbagai algoritma klasifikasi [24].

Secara keseluruhan, temuan-temuan ini menunjukkan bahwa model prediktif berdasarkan algoritma pembelajaran mesin klasik seperti Decision Tree, Naive Bayes, K-Nearest Neighbors, dan Support Vector Machine dapat memainkan peran penting dalam meningkatkan manajemen akademik dan mendukung intervensi dini bagi mahasiswa yang berisiko, meskipun masih diperlukan perbaikan, terutama dalam menangani dataset yang tidak seimbang yang melibatkan kasus keterlambatan kelulusan.

Penelitian ini bertujuan untuk memprediksi kelulusan tepat waktu mahasiswa menggunakan pendekatan pembelajaran mesin. Data dikumpulkan melalui kuesioner yang disebarakan kepada mahasiswa dan alumni Universitas Islam



Riau. Setelah proses pelabelan, data dipraproses melalui tahapan yang meliputi penanganan nilai yang hilang, normalisasi, dan pengkodean label. Untuk mengatasi ketidakseimbangan kelas antara mahasiswa yang lulus tepat waktu dan yang tidak, metode SMOTE (Synthetic Minority Over-sampling Technique) diterapkan.

Model kemudian dibangun menggunakan beberapa algoritma pembelajaran mesin klasik, yaitu Decision Tree, Gaussian Naive Bayes, K-Nearest Neighbors, dan Support Vector Machine (SVM). Setiap model dievaluasi dalam mengklasifikasikan mahasiswa ke dalam dua kategori: Lulus Tepat Waktu dan Tidak Lulus Tepat Waktu. Hasil klasifikasi diharapkan dapat mendukung pengambilan kebijakan akademik dan memberikan intervensi dini bagi mahasiswa yang berisiko mengalami keterlambatan kelulusan.

## **1.2. Perumusan Masalah**

Berdasarkan pada latar belakang diatas, maka dapat ditentukan beberapa rumusan pada penelitian ini, diantaranya sebagai berikut :

1. Bagaimana cara mengatasi ketidakseimbangan kelas dalam pembelajaran mesin untuk memprediksi kelulusan mahasiswa tepat waktu di Universitas Islam Riau?
2. Bagaimana mengembangkan model pembelajaran mesin yang akurat untuk memprediksi kelulusan mahasiswa tepat waktu di Universitas Islam Riau?
3. Bagaimana meningkatkan kinerja model pembelajaran mesin dalam memprediksi mahasiswa yang tidak lulus tepat waktu di Universitas Islam Riau?

## **1.3. Ruang Lingkup Penelitian**

Penelitian ini berfokus pada pengembangan model pembelajaran mesin untuk memprediksi kelulusan mahasiswa tepat waktu di Universitas Islam Riau dengan mengatasi masalah ketidakseimbangan kelas. Ruang lingkup penelitian ini meliputi:

1. Data yang digunakan adalah data mahasiswa Universitas Islam Riau yang meliputi data akademik, demografis, dan lain-lain.
2. Variabel yang digunakan adalah variabel yang relevan dengan kelulusan mahasiswa, seperti IPK, jumlah SKS, jenis kelamin, dan faktor lainnya yang berpengaruh.
3. Model pembelajaran mesin yang digunakan adalah model yang dapat menangani ketidakseimbangan kelas, seperti SMOTE, ADASYN, dan teknik lainnya yang relevan.
4. Evaluasi model dilakukan menggunakan metrik seperti akurasi, presisi, recall, dan F1-score.
5. Penelitian ini dilakukan di Universitas Islam Riau.
6. Penelitian ini dilakukan pada tahun 2025

#### **1.4. Tujuan Penelitian**

Tujuan yang hendak dicapai dalam pelaksanaan dari penelitian ini adalah :

1. Mengatasi masalah ketidakseimbangan kelas dalam pembelajaran mesin untuk memprediksi kelulusan mahasiswa tepat waktu di Universitas Islam Riau.
2. Mengembangkan model pembelajaran mesin yang akurat untuk memprediksi kelulusan mahasiswa tepat waktu di Universitas Islam Riau.
3. Meningkatkan kinerja model pembelajaran mesin dalam memprediksi mahasiswa yang tidak lulus tepat waktu di Universitas Islam Riau.

#### **1.5. Manfaat Penelitian**

Manfaat dari pelaksanaan penelitian ini adalah sebagai berikut :

1. Membantu UIR meningkatkan kualitas pendidikan dengan memprediksi kelulusan mahasiswa tepat waktu.
2. Membantu mahasiswa meningkatkan kemungkinan lulus tepat waktu.
3. Mengkontribusikan pada pengembangan model pembelajaran mesin yang akurat.

#### **1.6. Sistematika Penulisan**

Sistematika dalam penulisan laporan penelitian ini adalah sebagai berikut :

### **BAB I PENDAHULUAN**

BAB I ini berisikan Latar Belakang, Perumusan Masalah, Ruang Lingkup, Tujuan Penelitian, Manfaat Penelitian dan Sistematika Penulisan.

## **BAB II TINJAUANA PUSTAKA**

BAB II ini mencakup Tinjauan Pustaka dan Landasan Teori yang mendukung pelaksanaan sistem informasi Ketidakseimbangan Kelas dalam Pembelajaran Mesin untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu di Universitas Islam Riau

## **BAB III METODE PENELITIAN**

BAB III ini menjelaskan tahapan penelitian yang telah dilaksanakan, mulai dari literatur, persiapan penelitian beserta alat dan bahan yang dibutuhkan hingga metode analisa yang akan diterapkan.

## **BAB IV HASIL DAN PEMBAHASAN**

BAB IV ini menyajikan hasil dari rancangan sampai kepada mengujian sistem

## **BAB V KESIMPULAN DAN SARAN**

BAB V ini berisikan kesimpulan dan saran berdasarkan penelitian dan pembahasan yang telah di uraikan.

