

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Perguruan tinggi memiliki peran penting terhadap pengembangan sebuah negara (Hamsa, Indiradevi and Kizhakkethottam, 2016; Khan and Ghosh, 2021). Diantara berbagai tujuan, salah satunya adalah meningkatkan taraf kecerdasan suatu bangsa (Fadillah, 2015; Salah Hashim, Akeel Awadh and Khalaf Hamoud, 2020). Untuk mencapai tujuan tersebut, perguruan tinggi diharuskan memiliki sebuah sistem akademik. Sistem akademik mencakup seluruh proses pendidikan dari mulai penerimaan hingga tahap kelulusan seorang mahasiswa. Seluruh perguruan tinggi merancang sistem akademik sedemikian rupa untuk dapat menghasilkan mahasiswa yang berkualitas.

Setiap perguruan tinggi mempunyai caranya tersendiri untuk dapat bersaing dengan perguruan tinggi lainnya untuk menghasilkan lulusan terbaik. Kualitas mahasiswa yang dihasilkan oleh perguruan tinggi dipengaruhi oleh banyak faktor (Yang dan Li, 2018). Dalam memenuhi parameter kualitas luaran yang baik, diperlukan data historis terkait perkembangan kualitas pendidikan setiap mahasiswa yang ada (Puspita et al., 2018).

Diketahui dari penelitian yang telah dilakukan sebelumnya bahwa data historis proses pembelajaran dapat digunakan untuk memprediksi lama masa studi seorang mahasiswa (Andyana, 2015; Widhi Saputro and Wulan Sari, 2019). Penelitian sebelumnya juga telah berhasil membuat model yang dapat memprediksi nilai dari satu mata pelajaran yang diampu mahasiswa (Fadillah, 2015; Mayadewi and Rosely, 2015; Muflikhah, Yunita dan Furqon, 2017; Widaningsih, 2019; Kanetaki et al., 2022). Namun lama masa studi serta nilai satu mata kuliah hanya menjadi salah satu bagian dalam penilaian kualitas pendidikan maupun performa akademik seseorang secara keseluruhan.

Kemudian dari penelitian yang telah dilakukan, diketahui bahwa nilai akademik seperti Indeks Prestasi Kumulatif (IPK) dan nilai historis proses

pembelajaran dapat digunakan sebagai data untuk memprediksi performa mahasiswa (Ihsan Zul, 2013; Sravani and Bala, 2020; Bujang et al., 2021; Gufroni et al., 2021). Namun standarisasi IPK setiap universitas memiliki sistem perhitungan yang berbeda. Selain itu, prediksi menggunakan nilai progres akademik dapat disebut terlambat karena timbulnya permasalahan yakni gagalnya seseorang dalam suatu pembelajaran (Mishra, Kumar and Gupta, 2014; Imran et al., 2019).

Oleh karena itu diperlukan sebuah standar dengan skala lebih luas dan rentang waktu lebih dini untuk dapat membuat suatu sistem prediksi yang lebih relevan. Diketahui juga bahwa data admisi dapat digunakan sebagai data input untuk diproses oleh model kecerdasan buatan yang dapat mengukur performa akademik mahasiswa (Adekitan and Noma-Osaghae, 2018). Data admisi berupa berbagai macam data yang dikumpulkan untuk memasuki suatu perguruan tinggi. Penelitian tersebut berlanjut pada analisa data yang melihat seberapa besar pengaruh setiap komponen yang ada pada data admisi berupa Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) terhadap model untuk memprediksi performa akademik mahasiswa (Gufroni et al., 2021).

Pada tahun 2023, admisi yang berlaku untuk memasuki perguruan tinggi negeri di Indonesia terbagi menjadi 3 jalur. Diantaranya Seleksi Nasional Berdasarkan Prestasi (SNBP), Seleksi Nasional Berdasarkan Tes (SNBT) dan Jalur Mandiri (Said, Asbari and Salsabila, 2024). Namun pada 5 tahun sebelumnya, terdapat perbedaan penamaan diantaranya SNBP bernama Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), dan SNBT bernama Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN).

Penelitian ini akan mengolah data admisi perguruan tinggi berupa SNMPTN sebagai data percobaan. Data SNMPTN memiliki 25 fitur yang berisi identitas siswa, informasi asal sekolah, beserta 9 aspek penilaian dalam proses SNMPTN. Data tersebut akan diproses menjadi sebuah model yang dapat memprediksi performa akademik setiap mahasiswa. Data SNMPTN akan digabungkan dengan data nilai akademik mahasiswa untuk menguji apakah akan menghasilkan model yang optimal. Model pembelajaran mesin akan dibuat dengan membandingkan

algoritma *Supervised Machine Learning* Klasifikasi dengan nilai evaluasi yang paling tinggi. Parameter evaluasi yang umum digunakan untuk mengukur kualitas suatu model diantaranya nilai akurasi, *precision*, *recall*, dan *f1-score* (Luque et al., 2019). Terdapat beberapa algoritma yang umum digunakan untuk memecahkan kasus klasifikasi, diantaranya *Decision Tree* (DT), *Random Forest* (RF), dan *Support Vector Machine* (SVM) (Ul Hassan, Khan and Shah, 2018). Kemudian terdapat penelitian yang menunjukkan bahwa algoritma *Extreme Gradient Boosting* (XGBoost) dapat memberikan nilai evaluasi yang cukup baik pada kasus klasifikasi (Chen et al., 2018).

## 1.2. Rumusan Masalah

Berdasarkan uraian latar belakang diatas, diketahui terdapat beberapa permasalahan yakni :

1. Diperlukan sebuah pendekatan baru yang dapat mengolah data admisi untuk memprediksi performa akademik sehingga dapat dimanfaatkan sebagaimana mestinya. Pendekatan baru dapat dicoba dengan membangun sebuah model kecerdasan buatan.
2. Diperlukan uji coba untuk menemukan algoritma yang dapat menghasilkan nilai evaluasi tertinggi untuk prediksi performa akademik mahasiswa berdasarkan data admisi, dikarenakan model kecerdasan buatan pada data yang berbeda memerlukan algoritma yang sesuai untuk dapat menghasilkan nilai evaluasi terbaik.

## 1.3. Tujuan Penelitian

Tujuan yang ditargetkan dalam tugas akhir ini adalah :

1. Mengimplementasikan algoritma pembelajaran mesin yaitu DT, RF, SVM, dan XGB untuk membuat model yang dapat melakukan prediksi nilai mahasiswa dalam bentuk klasifikasi menggunakan data SNMPTN dan nilai akademik.
2. Mengevaluasi dan membandingkan model pembelajaran mesin prediksi performa akademik untuk mengetahui algoritma yang menghasilkan nilai evaluasi tertinggi.

#### 1.4. Batasan Masalah

Batasan masalah yang dibahas dalam tugas akhir ini adalah :

1. Data yang digunakan merupakan data admisi dari SNMPTN yang didapat dari suatu perguruan tinggi negeri di Indonesia dengan rentang tahun 2018 sampai dengan 2020 dan nilai IPK mahasiswa terkait pada 2 semester pertama.
2. Algoritma machine learning yang digunakan adalah Decision Tree, Random Forest, Support Vector Machine, dan XGBoost.
3. Pengujian model dilakukan dengan melihat parameter *accuracy*, *precision*, *recall*, dan *f1-score*.

#### 1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

1. Diketahui bahwa distribusi data pada data SNMPTN tergolong normal, dan rerata data nilai IPK banyak diisi pada rentang nilai  $3,00 < x \leq 4,00$ .
2. Dihasilkan model kecerdasan buatan yang dapat memprediksi performa akademik calon mahasiswa, yang dapat dijadikan rekomendasi pada proses penerimaan mahasiswa pada perguruan tinggi di Indonesia.
3. Model kecerdasan buatan yang dibuat dapat meningkatkan efisiensi proses penerimaan mahasiswa pada perguruan tinggi.
4. Dapat dilakukan pengecekan lebih lanjut terhadap model untuk menganalisa aspek dari pendidikan yang berperan penting pada hasil yang dikeluarkan model, yang dapat dijadikan landasan untuk meningkatkan sistem pendidikan pada perguruan tinggi.