

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pemanfaatan teknologi di bidang pendidikan telah meningkat dengan sangat pesat sehingga menjadikan setiap perguruan tinggi menghimpun data digital dengan jumlah yang sangat besar, baik itu data akademik, data kemahasiswaan, ataupun data dari layanan pendukung lainnya. Perguruan Tinggi memainkan peran penting dalam sebuah negara (Adekitan & Noma-Osaghae, 2019). Salah satu tujuan utamanya adalah untuk meningkatkan tingkat kecerdasan bangsa (Lemay et al., 2021). Untuk mencapai tujuan ini, Perguruan Tinggi harus memiliki Sistem Akademik mencakup seluruh proses pendidikan, mulai dari penerimaan mahasiswa baru, hingga kelulusan. Kualitas mahasiswa yang dihasilkan oleh institusi pendidikan tinggi dipengaruhi oleh banyak faktor (Fernandes et al., 2019), (Yang & Li, 2018). Untuk memenuhi parameter penilaian dengan hasil berkualitas baik, dibutuhkan data historis mengenai perkembangan kualitas pendidikan dari setiap mahasiswa yang ada (Romero & Ventura, 2020).

Institusi perguruan tinggi saat ini menghadapi tantangan dalam melakukan admisi seleksi penerimaan mahasiswa baru, dimana prosesnya harus dilaksanakan dengan cara yang benar untuk dapat memastikan kandidat calon mahasiswa yang diterima memiliki kemampuan yang tepat dalam memenuhi target akademik yang sesuai dengan bidang ilmu yang dipilihnya. Sering kali terjadi ketidaksesuaian kemampuan dari kandidat mahasiswa baru yang diterima dengan kemampuan yang diperlukan pada bidang yang dipilihnya menyebabkan hasil akademik tidak optimal, sehingga rentan terjadi potensi drop-out dari mahasiswa tersebut (Araque et al., 2009). Salah satu cara untuk menghindari hal tersebut adalah dengan melakukan prediksi kemampuan mahasiswa yang dilakukan sedini mungkin dimulai dari proses admisi seleksi penerimaan mahasiswa baru. Kemampuan

mahasiswa dapat diprediksi berdasarkan hasil dari ujian admisi, pada proses ini terkumpul data yang berkaitan dengan kemampuan peserta calon mahasiswa baru, data tersebut menjadi persyaratan penerimaan berdasarkan kriteria-kriteria dengan pembobotan yang telah ditentukan (Mengash, 2020). Selanjutnya dilakukan Prediksi mahasiswa berdasarkan kemampuan mahasiswa pada tahun ke-1 (semester 1 dan semester 2), hal ini dilakukan untuk melihat kesesuaian kemampuan mahasiswa dengan target akademik dari bidang keilmuannya. Perbandingan kemampuan mahasiswa pada proses admisi penerimaan mahasiswa baru, dengan kemampuan akademik mahasiswa di tahun pertama dapat dijadikan landasan dalam prediksi kemampuan akademik mahasiswa secara keseluruhan. Hal ini dapat mengidentifikasi mahasiswa dengan kemampuan rendah, dan kegagalan pada semester awal sehingga dapat dilakukan pencegahan agar terhindar dari drop-out (De Santos et al., 2019), (Nachouki & Naaj, 2022).

Penelitian sebelumnya menunjukkan bahwa data historis proses pembelajaran dapat digunakan untuk memprediksi lama masa studi mahasiswa (Alghazali et al., 2023), (Mengash, 2020), (Alshehri et al., 2020), dapat juga memprediksi nilai akhir dari matakuliah yang diambil mahasiswa (Adekitan et al., 2019), (A. Khan & Ghosh, 2021), (Kanetaki et al., 2022). Beberapa peneliti menggunakan Indeks Prestasi Kumulatif (IPK) atau *Grade Point Average* (GPA) untuk memprediksi kemampuan mahasiswa (Adekitan et al., 2019), (Sravani & Bala, 2020), (Bujang et al., 2021), (Gufroni et al., 2021), (Yaacob et al., 2019).

Prediksi kemampuan mahasiswa dapat dilakukan dengan pendekatan Educational Data Mining (EDM), EDM dapat diartikan sebagai proses data mining yang mengolah sejumlah dataset yang dihasilkan dari proses di lingkungan pendidikan untuk menyelesaikan permasalahan utama pada bidang pendidikan. EDM merupakan teknik yang sering digunakan untuk prediksi kemampuan mahasiswa (Aldowah et al., 2019), (Marbouti et al., 2016), (Romero & Ventura, 2020). EDM telah merevolusi prediksi kemampuan akademik dengan memungkinkan perguruan tinggi memproses dataset yang besar, dengan menggunakan *Machine Learning* (ML) berbasis *Supervised Learning Algorithm*,

seperti algoritma *Decision Tree* (DT), *Random Forest* (RF), *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB), para peneliti menunjukkan potensi besar EDM yang dapat digunakan untuk kepentingan perguruan tinggi dalam menganalisis dataset pendidikan dalam jumlah besar (De Santos et al., 2019).

Banyak penelitian menyoroti peran yang semakin besar dari ML dalam prediksi akademik. DT banyak digunakan karena kesederhanaannya dan kemudahan interpretasinya. Sebuah studi yang menggunakan DT menghasilkan tingkat akurasi lebih dari 80% dalam memprediksi mahasiswa berprestasi (Yaacob et al., 2019). Di sisi lain, RF secara konsisten memberikan akurasi prediksi lebih tinggi karena pendekatan ensemble learning-nya, dengan akurasi hingga 85,17% pada dataset penerimaan mahasiswa baru (Gufroni et al., 2021), RF digunakan untuk memprediksi nilai akhir mahasiswa dengan kondisi dataset yang tidak seimbang (*imbalanced*) dengan nilai F1-Score sebesar 99,5% (Bujang et al., 2021). SVM dikenal memiliki kelebihan dalam melakukan klasifikasi pada data yang berdimensi besar, penggunaan SVM untuk memprediksi kemampuan akademik mahasiswa dengan menggunakan IPK sebelumnya menghasilkan nilai akurasi sebesar 76,67%, lebih tinggi dari algoritma yang lainnya (Aluko et al., 2018). XGB juga muncul sebagai algoritma yang sangat efisien, dengan akurasi tertinggi mencapai 97,12% pada proses klasifikasi Prediksi hasil belajar sains siswa di sekolah (Harefa, 2024).

Berdasarkan penelitian sebelumnya, prediksi kemampuan akademik mahasiswa dapat dilakukan dengan menggunakan data nilai matakuliah (A. Khan & Ghosh, 2021), (Kanetaki et al., 2022), penggunaan IPK sebagai parameter untuk memprediksi kemampuan akademik mahasiswa (Sravani & Bala, 2020), (Bujang et al., 2021), (Gufroni et al., 2021), (Yaacob et al., 2019), selain menggunakan atribut data akademik, digunakan data pre-admisi pada ujian masuk sehingga menghasilkan model yang lebih akurat (Hassan & Al-Razgan, 2016). (Kanetaki et al., 2022) dalam penelitiannya menemukan bahwa kriteria persyaratan seleksi masuk bidang kognitif hanya memprediksi masing-masing 50,23% dari

kemampuan mahasiswa tahun pertama. Kemudian prediksi dengan atribut penilaian menggunakan data pre-admisi masih menjadi area yang kurang dieksplorasi pada bidang EDM (A. Khan & Ghosh, 2021).

Dengan membandingkan secara sistematis performa dari algoritma DT, RF, SVM, dan XGB, penelitian ini akan melakukan pengolahan data penerimaan mahasiswa baru dalam bentuk Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), yang saat ini berubah menjadi Seleksi Nasional Berbasis Prestasi (SNBP), data SNBP memiliki 25 fitur termasuk identitas mahasiswa, informasi asal sekolah, dan sembilan aspek penilaian dalam proses pre-admisi. Data SNBP ini digabungkan dengan data akademik mahasiswa pada tahun ke-1 (semester 1, dan semester 2), untuk diuji sehingga akan menghasilkan model yang memiliki nilai evaluasi tertinggi.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang diatas, maka rumusan masalah pada penelitian ini adalah:

1. Bagaimana melakukan proses pembobotan kriteria admisi berdasarkan data kemampuan mahasiswa pada saat SNBP, dan membandingkannya dengan kemampuan akademik mahasiswa pada tahun pertama perkuliahan?.
2. Bagaimana mengimplementasikan EDM dengan Supervised Learning Algorithm; DT, RF, SVM, dan XGB, dalam membuat model yang dapat melakukan prediksi Indeks Prestasi Kumulatif (IPK) mahasiswa menggunakan data pre-admisi dan kemampuan akademik IP semester 1, dan IP semester 2?.
3. Bagaimana mengevaluasi dan membandingkan model prediksi yang dibuat untuk mengetahui algoritma yang menghasilkan nilai evaluasi tertinggi?.

1.3. Tujuan Penelitian

Pada penelitian ini dapat diambil beberapa tujuan sebagai berikut:

1. Melakukan pembobotan kriteria admisi berdasarkan data kemampuan mahasiswa pada saat SNMPTN dan SNBP, dan membandingkan

kemampuan mahasiswa pada saat SNMPTN dan SNBP dengan kemampuan akademik mahasiswa pada tahun ke-1 perkuliahan (semester 1, dan semester 2).

2. Mengimplementasikan EDM dengan *Supervised Learning Algorithm*; DT, RF, SVM, dan XGB, dalam membuat model yang dapat melakukan prediksi Indeks Prestasi Kumulatif (IPK) mahasiswa menggunakan data pre-admisi dan nilai IP Semester 1, dan IP Semester 2.
3. Mengevaluasi dan membandingkan model prediksi kemampuan akademik untuk mengetahui algoritma yang menghasilkan nilai evaluasi tertinggi.

1.4. Manfaat Penelitian

Manfaat yang didapatkan pada penelitian ini adalah:

1. Penentuan pembobotan kriteria admisi SNMPTN dan SNBP dapat memudahkan Program Studi untuk menyesuaikan kriteria kemampuan mahasiswa yang akan diterimanya.
2. Menghasilkan model EDM untuk prediksi kemampuan akademik mahasiswa berbasis *Supervised Learning Algorithm* dengan nilai evaluasi yang tertinggi.
3. Model yang dihasilkan dapat menjadi acuan untuk Penerimaan Mahasiswa Baru (PMB) melalui jalur Mandiri pada periode selanjutnya sehingga akan mengurangi kemungkinan tingkat kegagalan mahasiswa (*drop-out*).

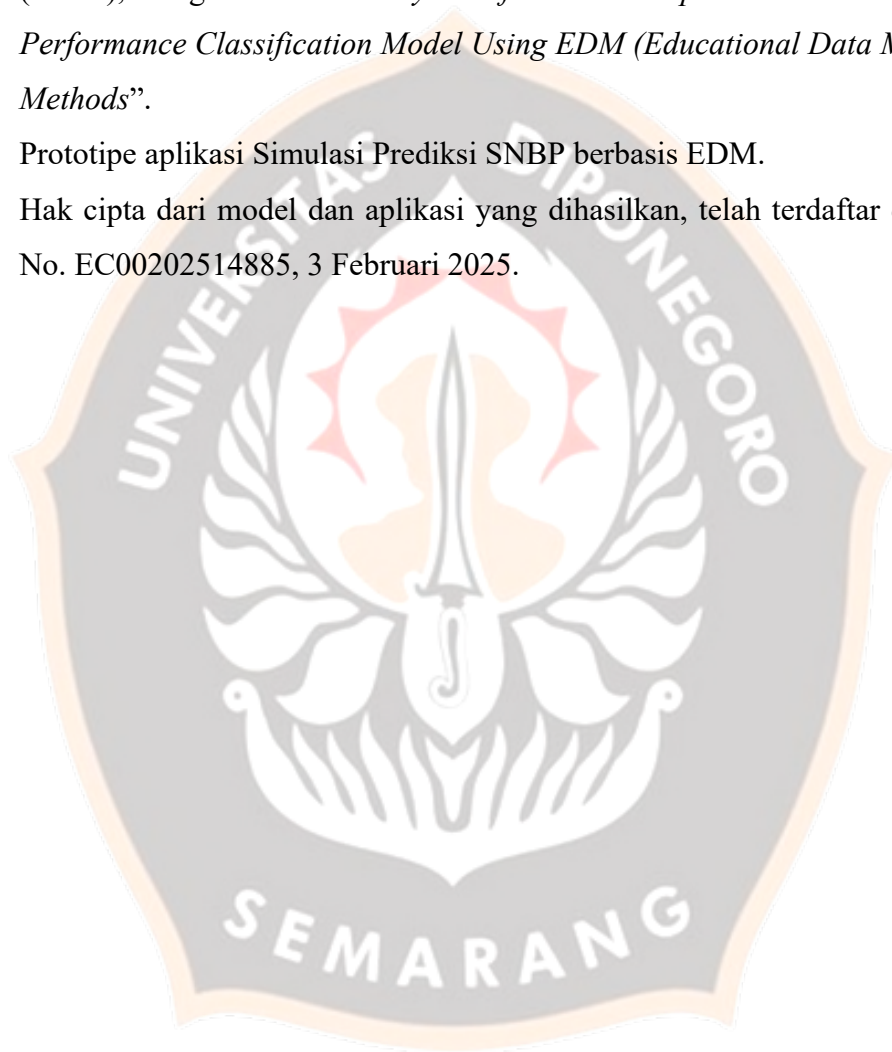
1.5. Luaran Penelitian

Luaran yang dihasilkan dari penelitian ini adalah:

1. Publikasi 1 paper pada seminar internasional terindeks Scopus dan IEEEExplore, telah terbit di ICICOS 2021 dengan judul "*Exploratory Data Analysis to Identify the Most Important Feature of University Admission Test Criteria Using Random Forest and Neural Network Algorithm*".
2. Publikasi 1 paper pada jurnal internasional terindeks Scopus, telah terbit di Journal on Informatics Visualization (JOIV) edisi Januari 2025, dengan

judul "*Academic Performance Prediction Using Supervised Learning Algorithms in University Admission*".

3. Publikasi 1 paper pada jurnal internasional terindeks Scopus, dengan status telah Accepted dan dalam proses revisi, pada Journal Applied Data Sciences (JADS), dengan Judul "*Analysis Of Feature Importance In Academic Performance Classification Model Using EDM (Educational Data Mining) Methods*".
4. Prototipe aplikasi Simulasi Prediksi SNBP berbasis EDM.
5. Hak cipta dari model dan aplikasi yang dihasilkan, telah terdaftar dengan No. EC00202514885, 3 Februari 2025.



SEKOLAH PASCASARJANA