

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Kualitas pendidikan tinggi sangat bergantung pada kemampuan institusi dalam memantau dan memprediksi performa akademik mahasiswanya secara akurat dan tepat waktu. Ketika penurunan prestasi dapat dideteksi lebih awal, pihak akademik memiliki kesempatan untuk melakukan intervensi yang tepat sasaran sebelum mahasiswa mengalami kegagalan akademik yang lebih serius [1]. Namun, proses identifikasi mahasiswa dengan potensi penurunan performa akademik secara konvensional masih banyak dilakukan melalui evaluasi manual berdasarkan rekam nilai, tingkat kehadiran, serta pengamatan terhadap perkembangan akademik mahasiswa. Pendekatan tersebut memiliki keterbatasan karena bergantung pada pengalaman evaluator dan sulit diterapkan secara efisien pada jumlah mahasiswa yang besar. Selain itu, metode konvensional cenderung bersifat reaktif karena identifikasi biasanya dilakukan setelah terjadi penurunan performa akademik.

Seiring dengan pesatnya perkembangan teknologi informasi, institusi pendidikan kini menghasilkan data dalam jumlah yang sangat besar, mulai dari nilai ujian, kehadiran, hingga aktivitas sosial dan ekonomi mahasiswa, yang berpotensi dimanfaatkan untuk keperluan analitik prediktif [2]. *Educational Data Mining* (EDM) merupakan bidang ilmu yang berfokus pada penerapan teknik-teknik penambangan data untuk mengekstraksi pengetahuan yang bermakna dari data pendidikan [3]. Dalam konteks EDM, prediksi performa akademik menjadi salah satu topik penelitian yang banyak dikaji karena relevansinya terhadap peningkatan kualitas pembelajaran dan pengambilan keputusan akademik. Pendekatan berbasis data memungkinkan institusi untuk menganalisis hubungan antara berbagai faktor akademik, sosial-ekonomi, dan perilaku mahasiswa secara lebih objektif dibandingkan pendekatan manual.

Berbagai algoritma *machine learning* telah diterapkan dalam domain EDM, di antaranya *Decision Tree*, *Naïve Bayes*, *Support Vector Machine* (SVM), *Random Forest*, hingga *Extreme Gradient Boosting* (XGBoost) [4]. Algoritma *ensemble learning* seperti *Random Forest* dan *XGBoost* banyak digunakan dalam penelitian prediksi performa akademik karena kemampuannya dalam menangani hubungan kompleks antar fitur serta menghasilkan performa prediksi yang kompetitif [5, 6,

7]. *Random Forest* dikenal memiliki kemampuan yang baik dalam mengurangi *overfitting* melalui kombinasi beberapa pohon keputusan, sedangkan *XGBoost* mampu meningkatkan performa prediksi melalui pendekatan *gradient boosting* [5, 6].

Meskipun berbagai algoritma *machine learning* telah diterapkan dalam prediksi performa akademik, penelitian sebelumnya masih memiliki beberapa keterbatasan. Sebagian penelitian berfokus pada penggunaan satu algoritma atau membandingkan performa model hanya berdasarkan nilai metrik evaluasi tanpa melakukan analisis lebih lanjut mengenai kontribusi masing-masing fitur terhadap hasil prediksi [2, 4]. Padahal, dalam konteks pendidikan, institusi tidak hanya membutuhkan model dengan performa prediksi yang baik, tetapi juga membutuhkan pemahaman mengenai faktor-faktor yang memengaruhi hasil prediksi agar informasi tersebut dapat dimanfaatkan dalam pengambilan keputusan akademik.

Berdasarkan celah penelitian tersebut, penelitian ini dilakukan untuk membandingkan secara langsung algoritma *Random Forest* dan *XGBoost* dalam memprediksi performa akademik mahasiswa. Selain mengevaluasi kemampuan prediksi kedua algoritma, penelitian ini juga melakukan analisis *feature importance* untuk mengidentifikasi fitur-fitur yang memiliki kontribusi terhadap hasil prediksi. Dengan demikian, penelitian ini diharapkan tidak hanya memberikan informasi mengenai performa model, tetapi juga memberikan gambaran mengenai faktor yang berpengaruh terhadap performa akademik mahasiswa.

1.2 Rumusan Masalah

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

1. Bagaimana perbandingan performa algoritma *Random Forest* dan *XGBoost* dalam memprediksi performa akademik mahasiswa berdasarkan metrik evaluasi klasifikasi?
2. Bagaimana mengukur fitur yang memiliki pengaruh besar dalam prediksi performa akademik mahasiswa berdasarkan analisis *feature importance*?

1.3 Batasan Permasalahan

Batasan permasalahan dalam penelitian ini adalah sebagai berikut.

1. *Dataset* yang digunakan adalah *Student Performance Metrics Dataset* yang diperoleh dari Mendeley Data dengan jumlah 493 data mahasiswa dan 16 atribut.
2. Penelitian ini hanya menggunakan algoritma *Random Forest* dan *XGBoost* sebagai model prediksi.
3. Variabel target berupa kategori performa akademik mahasiswa yang diperoleh melalui transformasi nilai *Overall*.
4. Atribut *Gender* dan *Hometown* tidak digunakan dalam proses pemodelan karena pertimbangan potensi bias dan karakteristik atribut.
5. Analisis faktor yang memengaruhi hasil prediksi dilakukan menggunakan nilai *feature importance* dari model *Random Forest* dan *XGBoost*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

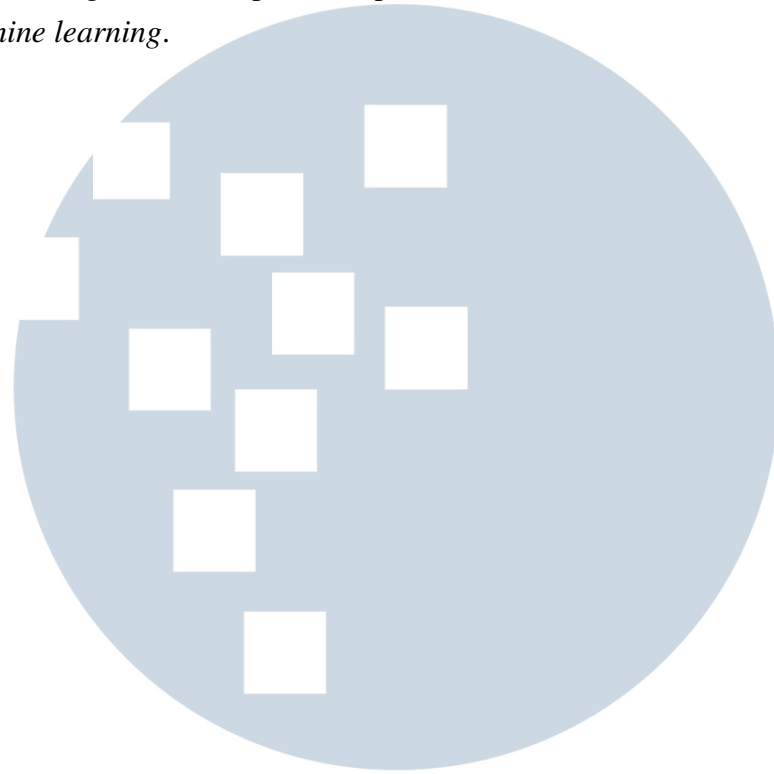
1. Membandingkan performa algoritma *Random Forest* dan *XGBoost* dalam memprediksi performa akademik mahasiswa menggunakan *Student Performance Metrics Dataset*.
2. Mengukur fitur yang memiliki pengaruh besar dalam prediksi performa akademik mahasiswa melalui analisis *feature importance*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut.

1. **Bagi Institusi Pendidikan:** Memberikan informasi mengenai algoritma *machine learning* dengan performa terbaik antara *Random Forest* dan *XGBoost* dalam prediksi performa akademik mahasiswa, serta memberikan gambaran mengenai faktor-faktor yang berpengaruh terhadap hasil prediksi sebagai pendukung pengambilan keputusan akademik.
2. **Bagi Peneliti:** Menambah kontribusi literatur di bidang *Educational Data Mining*, khususnya dalam perbandingan algoritma *ensemble learning* dan analisis faktor yang memengaruhi prediksi performa akademik mahasiswa.

3. **Bagi Pengembang Sistem:** Menyediakan referensi metodologis dalam pengembangan sistem prediksi performa akademik mahasiswa berbasis *machine learning*.



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA