

BAB I

PENDAHULUAN

1.1 Latar Belakang

Mengurangi angka putus sekolah siswa adalah salah satu tantangan yang dihadapi di sektor pendidikan secara global (Mduma et al., 2019). Berdasarkan data yang diperoleh dari UNESCO *Institute for Statistics* (UIS), secara global rerata persentase siswa yang putus pada sekolah menengah di tahun 2016 adalah 10,83% dan di tahun 2017 adalah 12,05%. Selain itu, angka mengulang/tidak naik kelas juga menjadi tantangan lain yang tidak bisa disepelekan, karena menurut estimasi UIS secara global terdapat 10 juta lebih siswa yang mengulang kelas (UNESCO *Institute for Statistics*, 2020). Tentunya angka tersebut cukup mengkhawatirkan, permasalahan ini sudah seharusnya menjadi perhatian utama bagi pelaku pendidikan (Aulck et al., 2016).

Salah satu faktor yang menyebabkan siswa mengulang bahkan putus sekolah yaitu terkendala secara akademik. Berdasarkan penelitian, gagal dalam pembelajaran di kelas serta tidak dapat mengikuti proses pembelajaran menjadi alasan terbesar mengapa siswa terdorong untuk berhenti sekolah (Doll et al., 2013). Maka dari itu, diperlukan suatu solusi yang dapat mengatasi permasalahan ini secara efektif, misalnya dengan intervensi dini pada siswa yang mengalami kendala akademik, karena hal tersebut sangat krusial untuk memastikan siswa dapat menyelesaikan sekolah tepat waktu (Mduma et al., 2019).

Keberhasilan dalam intervensi dini bergantung pada kemampuan sekolah untuk mengidentifikasi dan memprioritaskan siswa yang membutuhkan bantuan secara akurat (Lakkaraju et al., 2015), sehingga memerlukan teknik prediksi berbasis data yang dapat memfasilitasi penentuan siswa yang mengalami kendala akademik. Pendekatan menggunakan *Machine Learning* merupakan salah satu solusi yang patut dilirik untuk mengatasi hal ini (Mduma et al., 2019), karena dengan data akademik yang diperoleh, pendekatan menggunakan *Machine Learning* dapat mengidentifikasi siswa yang mengalami kendala akademik lebih dini dan akurat dibandingkan dengan pendekatan secara tradisional (Lakkaraju et al., 2015).

Berbagai penelitian yang telah dilakukan mengindikasikan potensi dari pendekatan *Machine Learning* menggunakan berbagai algoritma, misalnya *Decision Tree*, *C4.5*, *Naïve Bayes*, *Random Forest*, dan *SVM* (Ariyani, 2016; Beaulac & Rosenthal, 2019; Sugianto & Pujiyanita, 2019; Tumini & Damayanti, 2018; Zaenuri, 2019) untuk memprediksi kenaikan kelas ataupun kelulusan siswa. Penelitian-penelitian tersebut menghasilkan akurasi prediksi yang baik, yaitu diatas 80%. Namun penelitian-penelitian tersebut tidak membahas mengenai kelas minoritas, seperti siswa yang tidak naik kelas ataupun siswa yang tidak lulus.

Sedangkan pada penelitian (Lakkaraju et al., 2015) untuk identifikasi siswa beresiko tidak lulus tepat waktu (studi kasus dua distrik sekolah di U.S.) membandingkan beberapa algoritma, dengan hasil *Random Forest* unggul secara konsisten dibandingkan dengan algoritma lainnya. Tetapi penelitian ini hanya membandingkan beberapa algoritma *Machine Learning* tanpa fokus membahas bagaimana algoritma *Random Forest* dapat diimplementasikan untuk identifikasi siswa yang beresiko tidak lulus tepat waktu.

Maka dari itu, penulis melakukan penelitian dengan mengolah data siswa melalui implementasi Algoritma *Random Forest* untuk identifikasi dini siswa yang mengalami kendala akademik, sehingga dapat menjadi acuan bagi pelaku pendidikan yang berkaitan dalam melakukan intervensi lebih awal untuk meminimalisir angka mengulang kelas atau putus sekolah.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka rumusan masalah yang akan dibahas pada penelitian ini yaitu:

1. Bagaimana tahapan implementasi Algoritma *Random Forest* untuk identifikasi dini siswa sekolah menengah yang mengalami kendala akademik?
2. Apa saja atribut yang berpengaruh terhadap siswa sekolah menengah yang memiliki kendala akademik?
3. Bagaimana hasil implementasi Algoritma *Random Forest* untuk identifikasi dini siswa sekolah menengah yang mengalami kendala akademik?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu:

1. Mengetahui tahapan implementasi Algoritma *Random Forest* untuk identifikasi dini siswa sekolah menengah yang mengalami kendala akademik.
2. Mengetahui atribut apa saja yang berpengaruh terhadap siswa sekolah menengah yang memiliki kendala akademik.
3. Mengetahui hasil implementasi algoritma *Random Forest* untuk identifikasi siswa sekolah menengah yang mengalami kendala akademik.

1.4 Batasan Masalah

Berdasarkan masalah yang telah dirumuskan, agar pembahasan menjadi lebih terfokus diperlukan adanya pembatasan masalah. Berikut batasan masalah dalam penelitian ini.

1. Data yang digunakan adalah data sekunder, karena keterbatasan pengambilan data penelitian yang disebabkan oleh terjadinya Pandemi Covid-19.
2. Atribut yang diolah merupakan atribut yang ada pada data penelitian.
3. Proses pengolahan data menggunakan *library* scikit-learn Algoritma *Random Forest* yang terdapat pada perangkat lunak Anaconda Navigator.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini diantaranya:

1. Identifikasi dini siswa sekolah menengah yang mengalami kendala akademik merupakan langkah preventif bagi sekolah agar intervensi lebih awal dapat dilakukan.
2. Sebagai acuan bagi sekolah agar dapat mengambil tindak lanjut untuk memfasilitasi siswa sekolah menengah yang mengalami kendala akademik lebih dini, sehingga dapat meminimalisir angka tidak naik kelas ataupun ketidakiulusan siswa.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini dibagi menjadi 5 bab, dan masing-masing bab terdiri dari beberapa sub bab, yaitu:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Bab ini menguraikan materi yang berkaitan dengan penelitian. Materi yang disampaikan meliputi kendala akademik, *machine learning*, Algoritma *Random Forest*, serta *machine learning* dalam pendidikan.

BAB III METODE PENELITIAN

Bab ini memaparkan desain penelitian, objek penelitian, data penelitian sekunder yang digunakan, prosedur penelitian, serta tahapan analisis data yang digunakan dalam penelitian.

BAB IV PEMBAHASAN

Bab ini menganalisis hasil implementasi untuk identifikasi dini siswa yang terkendala akademik dengan Algoritma *Random Forest*, dari tahapan implementasi, *feature importance*, evaluasi model implementasi, serta hasil identifikasi siswa yang mengalami kendala akademik.

BAB V PENUTUP

Bab ini berisi kesimpulan dari keseluruhan penelitian yang telah dilakukan, implikasi, serta saran dari penulis untuk penelitian selanjutnya yang berkaitan dengan topik.