

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era Revolusi Industri 4.0, pentingnya analisis *data mining* meluas ke berbagai sektor, termasuk pendidikan (Triayudi & Widyarto, 2021). Munculnya *Educational data mining* (EDM) pada tahun 2005 menciptakan tren yang semakin mendapat perhatian, terutama di lembaga pendidikan global (Yağcı, 2022). Fokus utama saat ini adalah penerapan metode *data mining* di bidang pendidikan, khususnya dalam pengumpulan dan analisis data yang melibatkan mahasiswa, perkembangan studi, data dosen, daftar kompetensi lulusan, hasil kerja mahasiswa, ekstraksi pengetahuan, dan pengambilan keputusan pendidikan (Kovalev et al., 2020). Meskipun proses ini kompleks dan memerlukan upaya signifikan, EDM menjadi landasan penting dengan memproses berbagai jenis data, mengandalkan pola, asosiasi, dan hubungan antar data untuk mendapatkan informasi yang diinginkan (PJ et al., 2016). Dengan menggunakan teknik *data mining*, tujuan utama EDM adalah meningkatkan kualitas pendidikan dan menyelesaikan tantangan dalam dunia pendidikan (Alshareef et al., 2020). Selanjutnya, EDM juga bertujuan mengidentifikasi tren dan pola baru dari *big data* dengan menerapkan berbagai algoritma klasifikasi (Yağcı, 2022), dan di antara algoritma tersebut, Algoritma *Naïve Bayes* menjadi pilihan utama. Terkenal karena efektivitasnya dalam pengambilan keputusan klasifikasi, algoritma ini mampu mengantisipasi peluang masa depan berdasarkan pengalaman atau data di masa lalu (Hartatik et al., 2021). Dengan demikian, integrasi EDM dan Algoritma *Naïve Bayes* diharapkan memberikan kontribusi signifikan dalam mengatasi kompleksitas serta meningkatkan efektivitas keputusan dalam konteks pendidikan modern.

Studi terdahulu memfokuskan pada implementasi algoritma *Naïve Bayes* untuk memprediksi potensi mahasiswa *drop out* dan menganalisis alasan dibalik keputusan keluar. Faktor seperti kehadiran, riwayat pendidikan, riwayat medis, latar belakang keluarga, dan masalah disiplin dijadikan atribut dalam memprediksi risiko *drop out*

Karima Rizqia, 2024

PREDIKSI PREFERENSI ETIKA MAHASISWA MENGGUNAKAN ALGORITMA KLASIFIKASI NAÏVE BAYES
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

dari lembaga pendidikan tertentu (Suresh et al., 2017). Hasil penelitian di Universitas Dirgantara Marsekal Suryadarma menyoroiti akurasi tinggi algoritma *Naïve Bayes* dalam memprediksi kelulusan mahasiswa, dengan penekanan pada kecepatan dan efisiensi algoritma (Awaludin et al., 2022). Demikian pula, di Bulacan State University (BSU) Sarmiento Campus, penelitian menunjukkan bahwa algoritma klasifikasi *Naïve Bayes* efektif sebagai alat prediksi untuk menentukan probabilitas kelulusan mahasiswa, memberikan dasar bagi lembaga pendidikan untuk memberikan dukungan dan intervensi yang tepat kepada mahasiswa berisiko (Perez & Perez, 2021). Sementara itu, metode *Naïve Bayes* menjadi kritis dalam mengekstrak informasi mahasiswa yang memengaruhi faktor tertentu, seperti tingkat kelulusan, dengan data yang diperlukan mencakup NIM, nama mahasiswa, jenis kelamin, IP tahun pertama, IP tahun kedua, IPK, pilihan utama jurusan, pilihan kedua jurusan, dan riwayat pembayaran (Griha et al., 2022; Suganda et al., 2021). Selain itu, algoritma *Naïve Bayes* juga telah diterapkan untuk memprediksi performa mahasiswa dan aspek prediksi lainnya dalam konteks pendidikan (Farhana, 2021; Mustafa Saritas & Yasar, 2019; Pallathadka et al., 2023). Meskipun demikian, permasalahan umum dalam aplikasi *machine learning* dan *data mining* adalah ketidakseimbangan kelas, penelitian oleh (Thammasiri et al., 2014) menunjukkan bahwa teknik penyeimbangan kelas menggunakan *synthetic minority over-sampling technique* (SMOTE) berhasil mengatasi masalah klasifikasi yang tidak seimbang. Integrasi temuan ini memberikan kontribusi penting bagi lembaga pendidikan dalam meningkatkan efektivitas strategi pengelolaan mahasiswa.

Kajian mengenai penerapan algoritma *Naïve Bayes* dalam prediksi fenomena pendidikan telah menjadi fokus penelitian yang cukup luas di berbagai belahan dunia. Meskipun demikian, sebagian besar penelitian tersebut cenderung terfokus pada aspek-aspek akademik dan data informasi mahasiswa sebagai variabel utama. Penelitian yang lebih spesifik dan mendalam mengenai prediksi preferensi etika mahasiswa, yang mencakup eksplorasi perilaku etis mahasiswa dalam lingkup pendidikan, masih relatif kurang terbahas. Maka penelitian ini akan fokus membahas pada implementasi algoritma klasifikasi *Naïve Bayes* dalam memprediksi preferensi etika mahasiswa.

Fokusnya tidak hanya mencakup aspek akademik, tetapi juga merinci beberapa faktor yang dianggap memiliki dampak signifikan pada perilaku etis mahasiswa. Beberapa faktor yang menjadi fokus penelitian antara lain motivasi belajar, efikasi diri, *resilience*, *team strain*, *knowledge articulation*, dan *cooperative classroom environment*. Dengan mempertimbangkan faktor-faktor ini, penelitian ini bertujuan untuk memprediksi preferensi etika mahasiswa dan menyelidiki hubungan antara atribut-atribut tersebut dengan tingkat etika mahasiswa. Proses analisis data dalam penelitian ini akan melibatkan implementasi aplikasi *RapidMiner Studio*. *RapidMiner Studio* adalah salah satu perangkat lunak *data mining* yang terkenal (Madyatmadja et al., 2021) juga sebagai perangkat lunak yang telah terbukti, diharapkan dapat menyederhanakan dan meningkatkan efisiensi analisis data, terutama dalam menghadapi volume data besar atau *big data* (Uysal et al., 2022). Pentingnya penelitian ini tidak hanya terletak pada kontribusinya terhadap pemahaman perilaku etis mahasiswa dan manfaatnya untuk lembaga pendidikan tinggi, tetapi juga pada penerapan teknologi seperti *RapidMiner Studio* sebagai alat yang dapat membuka peluang untuk eksplorasi lebih lanjut dan perluasan penelitian di bidang ini. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap literatur dan praktik pendidikan dalam merinci dan memahami faktor-faktor yang memengaruhi model etika mahasiswa.

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah:

1. Bagaimana perbandingan hasil penggunaan algoritma *Naïve Bayes* untuk klasifikasi menggunakan analisis korelasi dan SMOTE dalam memprediksi preferensi etika mahasiswa?
2. Bagaimana perbandingan tingkat akurasi prediksi algoritma *Naïve Bayes* tanpa dan dengan menggunakan analisis korelasi dan SMOTE dalam memprediksi preferensi etika mahasiswa?

3. Apa saja faktor-faktor yang signifikan dan mempengaruhi prediksi preferensi etika mahasiswa berdasarkan analisis korelasi?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah:

1. Mengimplementasikan aplikasi *machine learning* yaitu algoritma klasifikasi *Naïve Bayes* untuk prediksi preferensi etika mahasiswa.
2. Menghasilkan informasi atau rekomendasi untuk lembaga pendidikan tinggi mengenai faktor-faktor yang mempengaruhi etika mahasiswa.
3. Memanfaatkan aplikasi *RapidMiner Studio* untuk memastikan efisiensi *educational data mining*.

1.4 Batasan Penelitian

Untuk menjamin keakuratan dan relevansi hasil penelitian, penting untuk menetapkan batasan masalah. Dalam konteks penelitian ini, berikut adalah cakupan dan batasan yang diterapkan:

1. Penelitian ini menggunakan data sekunder yang diambil dari situs web *Data in Brief* www.sciencedirect.com yang telah *open acces* Data yang digunakan adalah data model etika mahasiswa perguruan tinggi, data ini dikumpulkan antara Juli 2018 sampai dengan Desember 2018 di Yogyakarta, Indonesia.
2. Penelitian ini menggunakan algoritma klasifikasi *Naïve Bayes* serta menerapkan analisis korelasi dan metode SMOTE untuk pemodelan prediksi.
3. Atribut yang digunakan untuk menentukan prediksi preferensi etika mahasiswa yaitu motivasi belajar, efikasi diri, *resilience*, *knowledge articulation*, *team strain*, dan *cooperative classroom environment*.

4. Hasil dari prediksi yang akan dihasilkan mencakup data tingkat etika mahasiswa dan evaluasi kinerja model algoritma Naïve Bayes dalam memprediksi preferensi etika mahasiswa.
5. Alat bantu untuk pengujian yang dipakai adalah *RapidMiner Studio*.

1.5 Manfaat Penelitian

Manfaat secara teoritis pada penelitian ini diharapkan dapat menjadi referensi untuk pembaca dan peneliti untuk memperluas wawasan mengenai implementasi algoritma *Naïve Bayes* dan memahami perilaku etis mahasiswa dalam konteks pendidikan, adapun secara praktis penelitian ini bisa bermanfaat bagi lembaga pendidikan dan peneliti selanjutnya yang membahas pengembangan strategi pendidikan yang lebih efektif dalam meningkatkan perilaku etis mahasiswa dan dapat lebih mendalam terfokus pada aspek-aspek tertentu yang muncul dari hasil penelitian ini.

1.6 Struktur Organisasi Skripsi

Sistematika penulisan skripsi terdapat lima bab yang merujuk pada pedoman penulisan karya ilmiah Universitas Pendidikan Indonesia tahun 2019. Pada Bab I terdiri dari latar belakang, rumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian dan struktur organisasi skripsi. Bab II terdiri dari tinjauan pustaka mengenai teori yang berhubungan dengan *educational data mining*, algoritma klasifikasi *Naïve Bayes*, perilaku etika mahasiswa, aplikasi *RapidMiner Studio*, *machine learning*, dan ketidakseimbangan kelas (*imbalanced class*). Bab III membahas metode penelitian mencakup desain penelitian, sumber data, partisipan, instrumen penelitian, prosedur penelitian, serta pengolahan dan analisis data. Bab IV menjelaskan hasil dan pembahasan dari data yang telah diolah untuk menjawab pertanyaan penelitian yang telah dirumuskan. Bab V terdiri dari simpulan penelitian, implikasi penelitian, dan rekomendasi berdasarkan hasil penelitian yang telah dilakukan.