

BAB I

PENDAHULUAN

Strategi sumber daya manusia yang kuat bergantung pada fondasi pendidikan yang kuat untuk menumbuhkan tenaga kerja yang kompetitif, menggarisbawahi peran penting sistem pendidikan modern dalam mendorong perkembangan siswa. Inti dari tujuan ini adalah penerapan metodologi penilaian yang efektif, yang berfungsi sebagai alat untuk mengevaluasi pemahaman siswa, menginformasikan keputusan kurikuler, dan menyempurnakan praktik instruksional. Penilaian, yang didefinisikan sebagai “perolehan informasi secara sistematis untuk memandu keputusan mengenai siswa, kurikulum, metode pengajaran, atau kebijakan pendidikan” [1], sangat penting dalam memastikan akuntabilitas dan peningkatan berkelanjutan dalam lembaga pendidikan. Studi ini bertujuan untuk meningkatkan kualitas pendidikan di Sekolah Menengah Pertama (SMP) dan Madrasah Tsanawiyah (MTs) di Indonesia dengan mengatasi kesenjangan dalam praktik penilaian dan pemanfaatan data yang ada.

Untuk mencapai tujuan ini, sangat penting untuk mengidentifikasi dan menganalisis berbagai faktor yang mempengaruhi kinerja siswa. Sekolah secara rutin mengumpulkan beragam data - termasuk catatan akademik, pola kehadiran, keterlibatan ekstrakurikuler, latar belakang sosial ekonomi, dan akses ke fasilitas pembelajaran - yang memiliki potensi yang belum dimanfaatkan untuk memahami faktor-faktor penentu keberhasilan akademik[2]-[4]. Namun, ketergantungan pada pemrosesan data manual dan metode analisis tradisional telah menyebabkan kurangnya pemanfaatan sumber daya ini secara signifikan, sehingga membatasi kemampuan pendidik untuk mendapatkan wawasan yang dapat ditindaklanjuti[5]-[7]. Dengan memanfaatkan pendekatan berbasis data yang canggih, penelitian ini berupaya mengubah data pendidikan mentah menjadi pengetahuan strategis, sehingga memungkinkan para pemangku kepentingan untuk merancang intervensi yang ditargetkan untuk mengatasi ketidaksetaraan sistemik dan mengoptimalkan hasil pembelajaran.

Dalam pemodelan prediktif, [8], [9] masih menjadi tantangan yang kritis, terutama pada dataset di mana distribusi variabel target sangat miring. Masalah ini sering kali membuat algoritma konvensional memprioritaskan kelas mayoritas dan mengabaikan kelas minoritas (misalnya, siswa yang berisiko putus sekolah), sehingga menghasilkan akurasi prediksi yang bias dan tidak realistis [10]-[12]. Untuk mengatasi keterbatasan ini, teknik pembelajaran

ketidakseimbangan telah dikembangkan untuk meningkatkan kemampuan model prediktif dalam mengenali kelas-kelas yang kurang terwakili secara efektif. Teknik-teknik ini beroperasi melalui dua strategi utama: memodifikasi algoritme prediksi untuk memperhitungkan ketidakseimbangan kelas secara inheren atau meningkatkan kualitas representasi dataset itu sendiri.

Sebagai contoh, metode preprocessing seperti oversampling dapat digunakan untuk menyeimbangkan proporsi kelas dengan memperkuat sampel minoritas secara sintesis, sehingga menyempurnakan data input untuk model hilir [8], [13]-[15]. Sebagai alternatif, penyesuaian algoritmik, seperti pengklasifikasi berbasis ensemble (misalnya, metode agregasi bootstrap), dapat diimplementasikan untuk secara iteratif memprioritaskan contoh minoritas yang salah diklasifikasikan selama pelatihan. Dalam penelitian ini, istilah “teknik pembelajaran ketidakseimbangan” digunakan secara bergantian dengan “teknik resampling,” karena fokusnya terutama terletak pada peningkatan yang berpusat pada data daripada modifikasi arsitektur model. Dengan memprioritaskan penyempurnaan dataset, pendekatan ini memastikan bahwa model prediktif mencapai kinerja yang lebih adil dan dapat digeneralisasi di semua kelas.

Meskipun terdapat kemajuan yang signifikan dalam teknik pembelajaran ketidakseimbangan, penelitian yang ada saat ini lebih memprioritaskan pengoptimalan algoritmik (misalnya, menyempurnakan mekanisme teknis metode seperti SMOTE atau pengklasifikasi ansambel) daripada aplikasi praktis dalam konteks dunia nyata [16], [17]. Meskipun berbagai penelitian telah mengeksplorasi peningkatan teknis—seperti meningkatkan algoritme pengambilan sampel atau metode hibridisasi—masih ada kesenjangan yang kritis. Pertama, banyak evaluasi yang bergantung pada set data simulasi, sehingga membatasi wawasan tentang bagaimana teknik ini bekerja di bawah batasan data dunia nyata. Kedua, ketika diterapkan, strategi pembelajaran ketidakseimbangan sering kali diperlakukan sebagai alat bantu untuk pengembangan model, tanpa diskusi yang memadai tentang tantangan implementasi, interpretabilitas, atau kesesuaian kontekstualnya.

Penelitian ini bertujuan untuk meningkatkan akurasi model prediktif dalam mengidentifikasi siswa yang berisiko (misalnya, mereka yang berisiko putus sekolah atau gagal mendaftar di pendidikan pascasekolah menengah) di sekolah menengah atas dengan mengatasi ketidakseimbangan kelas dalam set data pendidikan dunia nyata [12], [17]. Dataset yang representatif secara nasional yang terdiri dari variabel-variabel kompleks seperti kinerja akademik, keterlibatan ekstrakurikuler, latar belakang sosial ekonomi, dan akses ke sumber daya

pembelajaran-kami menerapkan SMOTE (Synthetic Minority Over-sampling Technique) untuk memitigasi rasio ketimpangan yang parah. SMOTE diprioritaskan karena kemampuannya untuk menghasilkan sampel sintesis yang secara statistik mereplikasi karakteristik kelas yang kurang terwakili, sehingga meningkatkan sensitivitas model terhadap pola minoritas [10], [17], [18].

Untuk memastikan prediksi yang kuat, tiga algoritma klasifikasi - Decision Tree (kemampuan interpretasi yang tinggi)[19]-[21], Random Forest (ketahanan terhadap interaksi data yang kompleks)[16], [22], [23], dan SVM (Support Vector Machine) (keefektifan di ruang dimensi tinggi) - dioptimalkan dan dievaluasi secara sistematis[22], [24], [25]. Model Decision Tree digunakan untuk mengidentifikasi faktor risiko kritis secara transparan (misalnya, tingkat kehadiran, keterlibatan orang tua), sementara Random Forest dan SVM diuji untuk mengetahui kemampuannya dalam menangani hubungan nonlinier dan noise yang melekat pada set data. Metrik kinerja yang menekankan pada hasil kelas minoritas (presisi, recall, F1-score) digunakan untuk mengevaluasi kemampuan model, memastikan keadilan dalam prediksi untuk subkelompok siswa yang rentan.

Implementasi Langsung di Sekolah Menengah Pertama Terlepas dari kumpulan data, metodologi ini dirancang untuk memberikan wawasan yang dapat ditindaklanjuti bagi para pemangku kepentingan sekolah menengah Pertama. Dengan menyelaraskan variabel prediktor dengan kerangka kerja pendidikan yang sudah ada, seperti Model Putus Sekolah dari Tinto (kegigihan akademik) dan Teori Sistem Ekologi Bronfenbrenner (pengaruh lingkungan), fitur-fitur seperti konsistensi kehadiran, pendapatan rumah tangga, dan akses ke bimbingan belajar dimasukkan ke dalam model.