

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Pendidikan menjadi fondasi utama dalam pengembangan sumber daya manusia yang berkualitas. Di tingkat SMP, motivasi belajar dan prestasi akademik berperan besar dalam keberhasilan proses pembelajaran siswa, terutama pada masa perkembangan psikologis yang dinamis pada usia ini. Berbagai faktor internal dan eksternal seperti kehadiran, waktu belajar, partisipasi, dan dukungan lingkungan memiliki efek signifikan terhadap motivasi dan hasil belajar siswa [1], [2].

Revolusi digital telah membuka peluang baru dengan penerapan machine learning dalam ranah pendidikan. Berbagai studi menunjukkan bahwa algoritma ML, terutama *Random Forest* (RF) dan *Logistic Regression* (LR), mampu mendeteksi pola tersembunyi dalam data akademik dan memberikan prediksi yang akurat [3], [4]. RF, sebagai metode ensemble, sering mencapai akurasi tinggi dalam klasifikasi akademik dan mampu mengatasi variabel dengan korelasi kuat tanpa resiko overfitting [5], [6]. Sementara itu, LR tetap diandalkan karena interpretabilitas sederhana dan performa baik untuk modeled klasifikasi linier [7].

Seiring kemajuan teknologi informasi, pendekatan analisis pendidikan tidak lagi hanya bergantung pada metode konvensional. Kini, *machine learning* (ML) hadir sebagai alternatif modern yang mampu menganalisis data pendidikan dalam skala besar dan menemukan pola-pola tersembunyi yang sulit terdeteksi secara manual [3]. Dalam beberapa tahun terakhir, penelitian mengenai penerapan ML dalam dunia pendidikan meningkat secara signifikan, baik untuk memprediksi kelulusan, prestasi akademik, maupun tingkat motivasi belajar siswa [2], [8].

Dua algoritma ML yang paling banyak digunakan dalam klasifikasi dan prediksi dalam konteks pendidikan adalah *Random Forest* (RF) dan *Logistic Regression* (LR). Random Forest dikenal sebagai metode ensemble yang kuat karena menggabungkan banyak pohon keputusan (*decision tree*) untuk mencapai akurasi tinggi dan mengurangi risiko overfitting [3], [5]. RF juga mampu memberikan informasi penting mengenai feature importance, sehingga sangat berguna dalam mengidentifikasi variabel-variabel yang paling berpengaruh terhadap motivasi dan prestasi siswa [6].

Sementara itu, Logistic Regression adalah metode klasik namun tetap relevan dalam berbagai kasus klasifikasi, terutama ketika hubungan antar variabel bersifat linier dan interpretasi model menjadi prioritas [2], [7]. Kelebihan utama dari LR terletak pada kemampuannya memberikan pemodelan yang sederhana namun cukup kuat dalam memprediksi kemungkinan kejadian seperti motivasi rendah atau tinggi, serta prediksi kelulusan siswa.

Beberapa studi telah membandingkan performa kedua metode ini dalam konteks prediksi akademik. Misalnya, penelitian oleh Koper [6] menyebutkan bahwa RF menghasilkan akurasi yang lebih tinggi dibandingkan LR dalam prediksi hasil ujian matematika siswa. Namun, LR dinilai lebih mudah diinterpretasi dan lebih cepat dalam proses pelatihan model [7]. Ini menunjukkan adanya trade-off antara akurasi dan interpretabilitas, yang menjadikan perbandingan kedua metode ini sangat relevan untuk dikaji lebih lanjut dalam konteks prediksi motivasi dan hasil belajar siswa SMP.

Penelitian di tingkat SMP dan SMA di Indonesia pun menunjukkan keberhasilan penerapan RF untuk memprediksi prestasi dan motivasi akademik [8], [9]. Namun, ada kebutuhan untuk membandingkan secara langsung performa RF dan LR dalam ranah SMP, terutama apakah RF lebih unggul dalam akurasi atau LR menawarkan kelebihan interpretasi praktis [10].

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan metode Random Forest dan Logistic Regression dalam memprediksi motivasi dan prestasi akademik siswa SMP, dengan harapan dapat memberikan rekomendasi model yang seimbang antara akurasi dan kegunaan praktis di lingkungan pendidikan dasar di Indonesia.

1.2. Identifikasi Masalah

Masalah yang teridentifikasi:

1. Kurangnya pemanfaatan data dalam memetakan motivasi dan prestasi belajar siswa SMP.
2. Belum banyak penelitian yang membandingkan algoritma ML di tingkat SMP.
3. Dibutuhkan metode prediktif yang akurat dan mudah dipahami oleh pendidik.
4. Minimnya penerapan machine learning sebagai alat bantu keputusan di sekolah dasar-menengah.
5. Belum diketahui algoritma yang lebih optimal antara Random Forest dan Logistic Regression dalam konteks ini.

1.3. Batasan Masalah

Agar penelitian ini lebih terfokus, maka batasan masalah dalam penelitian ini adalah:

1. Penelitian ini hanya berfokus pada prediksi motivasi dan hasil belajar siswa Sekolah Menengah Pertama (SMP), dan tidak mencakup jenjang pendidikan lain seperti SD atau SMA.
2. Data yang digunakan adalah dataset terbuka (open dataset) mengenai data pendidikan siswa yang tersedia di platform terpercaya seperti Kaggle, yang mencakup informasi terkait motivasi belajar, prestasi akademik, serta faktor-faktor pendukung lainnya.
3. Penelitian ini hanya memfokuskan penggunaan dua algoritma machine learning, yaitu Random Forest dan Logistic Regression, tanpa membandingkan dengan metode lainnya seperti SVM, KNN, atau XGBoost.
4. Evaluasi performa model dibatasi pada empat metrik evaluasi utama, yaitu: akurasi, sensitivitas (recall), spesifisitas, dan AUC.

1.4. Rumusan Masalah

Berdasarkan latar belakang di atas, terdapat rumusan masalah yaitu :

1. Untuk mengetahui apa perbedaan kinerja algoritma Random Forest dan Logistic Regression dalam memprediksi motivasi dan hasil belajar siswa SMP, serta bagaimana proses pengujiannya?
2. Untuk mengetahui bagaimana tingkat akurasi, sensitivitas, spesifisitas, dan AUC dari Random Forest dan Logistic Regression dalam memprediksi motivasi dan hasil belajar siswa SMP?
3. Untuk mengetahui metode apa yang paling optimal dan mengapa metode tersebut lebih unggul dalam mendukung pengambilan keputusan pendidikan di SMP?

1.5. Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

1. Untuk mengetahui apa perbedaan kinerja algoritma Random Forest dan Logistic Regression dalam memprediksi motivasi dan hasil belajar siswa SMP.
2. Untuk mengetahui bagaimana performa algoritma Random Forest dan Logistic Regression berdasarkan akurasi, sensitivitas, spesifisitas, dan AUC.
3. Untuk mengetahui metode apa yang paling optimal dan bagaimana penerapannya sebagai pendukung pengambilan keputusan di bidang pendidikan SMP.

1.6. Manfaat Penelitian

1.6.1 Bagi Peneliti

Peneliti mendapatkan pemahaman empiris mengenai kelebihan dan keterbatasan algoritma Random Forest dan Logistic Regression dalam konteks prediksi motivasi dan hasil belajar siswa. Selain itu, penelitian ini dapat menjadi landasan awal bagi studi lanjutan yang lebih kompleks dengan metode ML lainnya atau integrasi dengan pendekatan psikometrik.

1.6.2 Bagi Siswa/Siswi

Penelitian ini memberikan gambaran konkret mengenai faktor-faktor kunci yang berpengaruh terhadap motivasi dan prestasi belajar, seperti waktu belajar, tingkat kehadiran, partisipasi aktif, dan dukungan lingkungan belajar. Dengan informasi ini, siswa/siswi dapat lebih sadar akan kebiasaan dan perilaku belajar yang harus ditingkatkan guna meraih hasil akademik yang lebih baik serta membangun motivasi internal yang berkelanjutan.

1.6.3 Bagi Tenaga Pendidik (Guru)

Penelitian ini menyediakan alat bantu analisis prediktif yang dapat digunakan oleh guru untuk mengidentifikasi siswa dengan potensi risiko rendahnya motivasi atau pencapaian belajar. Dengan demikian, guru dapat merancang strategi intervensi yang lebih tepat sasaran, seperti pemberian

bimbingan tabahan, pendekatan personal dalam pembelajaran, atau dukungan emosional yang disesuaikan dengan karakteristik siswa.

1.6.4 Bagi Instansi Sekolah

Hasil penelitian ini dapat dimanfaatkan oleh pihak sekolah sebagai dasar dalam penyusunan kebijakan pembelajaran berbasis data, serta mendukung program peningkatan mutu pendidikan. Sekolah dapat menggunakan hasil prediksi untuk mengalokasikan sumber daya secara lebih efisien, menyusun program mentoring, serta mengevaluasi efektivitas strategi dan kurikulum yang telah diterapkan dalam meningkatkan motivasi dan prestasi akademik siswa.