BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pendidikan adalah upaya untuk membentuk manusia yang berperikemanusiaan, sebuah usaha untuk meningkatkan martabat dan kehormatan manusia sehingga mampu saling menghormati satu sama lain. Pendidikan menjadi pilar utama dalam kemajuan suatu bangsa dan negara. Salah satu indikator kemajuan sebuah bangsa dan negara adalah kualitas serta sistem pendidikan yang diterapkan (Umadi, 2023).

Pemerataan pendidikan sangat penting untuk diwujudkan. Namun, kenyataannya masih ada anak-anak yang tidak melanjutkan sekolah (Azkiyah et al., 2025). Data Angka Partisipasi Sekolah (APS) yang tersedia menunjukkan bahwa pemerataan pendidikan masih kurang, karena hingga saat ini angkanya belum mencapai 100% (Zulfa Rahmatin & Ady Soejoto, 2017). Salah satu indikator capaian pendidikan adalah APS. APS adalah perbandingan antara jumlah penduduk kelompok usia sekolah yang bersekolah terhadap seluruh penduduk kelompok usia sekolah (Rumbiak et al., 2023). Semakin tinggi nilai APS, semakin berhasil suatu daerah dalam memberikan akses ke layanan pendidikan. APS yang tinggi menandakan peluang yang lebih besar untuk akses pendidikan secara umum. Peluang ini dapat diamati pada kelompok umur tertentu melalui besaran APS masing-masing kelompok umur tersebut (Safitri et al., 2024).

Kemajuan teknologi informasi dan komunikasi telah memberikan dampak besar dalam berbagai aspek kehidupan, termasuk pendidikan (Huraerah et al., 2023). Salah satu teknologi yang berkembang pesat adalah algoritma *Machine Learning*. Algoritma ini memungkinkan komputer untuk belajar dari data tanpa perlu pemrograman eksplisit. Dengan kemampuannya memprediksi berdasarkan pola dalam data, algoritma *Machine Learning* memiliki potensi besar untuk diterapkan di berbagai bidang, termasuk memprediksi angka partisipasi sekolah di Indonesia (Oktavianus et al., 2023). Prediksi APS dapat

memberikan informasi penting bagi para pembuat kebijakan dan institusi pendidikan dalam merencanakan dan mengambil keputusan terkait pengalokasian sumber daya dan peningkatan kualitas pendidikan di masa yang akan datang. Dengan memprediksi APS, pemerintah dan institusi pendidikan dapat mengetahui berapa banyak siswa yang diharapkan untuk masuk ke dalam sistem pendidikan pada suatu periode waktu tertentu, sehingga mereka dapat mengatur dan menyesuaikan sumber daya dan kebijakan pendidikan yang dibutuhkan untuk memenuhi kebutuhan tersebut. Selain itu, prediksi APS juga dapat membantu pemerintah dan institusi pendidikan untuk memperkirakan kebutuhan tenaga pengajar, buku, dan bahan ajar, serta fasilitas pendukung lainnya yang diperlukan untuk menunjang proses belajar mengajar. Hal ini dapat membantu meningkatkan efisiensi dan efektivitas system pendidikan secara keseluruhan (Aji et al., 2023).

Namun dalam upaya memprediksi APS menggunakan algoritma *Machine Learning* terdapat beberapa tantangan seperti kurangnya referensi jurnal yang memprediksi APS menggunakan algoritma *Machine Learning*, serta pemilihan algoritma yang tepat. Dengan memilih algoritma yang sesuai, maka hasil yang lebih akurat akan dapat dicapai (Trisna et al., 2025). Seperti yang dilakukan oleh (Wanto, 2019) dalam jurnal “Prediksi Angka Partisipasi Sekolah dengan Fungsi Pelatihan *Gradient Descent with Momentum & Adaptive LR*” menunjukkan bahwa dalam memprediksi APS dengan menggunakan algoritma *Linear Regression* berhasil mencapai tingkat akurasi sebesar 91% dan MSE 0,0023428551. Penelitian yang mendekati dengan penelitian sebelumnya yaitu penelitian oleh (Kharis et al., 2023) dalam jurnal “Prediksi Kelulusan Siswa pada Mata Pelajaran Matematika menggunakan *Educational Data Mining*” menunjukkan bahwa dalam memprediksi kelulusan siswa dengan menggunakan algoritma *XGBoost* berhasil mencapai tingkat akurasi sebesar 90,89%. Selain itu, penelitian oleh (Zeniarta et al., 2022) dalam jurnal “Seleksi Fitur dan Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa” menunjukkan bahwa dalam memprediksi kelulusan mahasiswa dengan menggunakan algoritma *Random Forest* berhasil mencapai tingkat akurasi sebesar 77,35%.

Penelitian ini akan secara khusus membandingkan kinerja dari 3 algoritma *Machine Learning*, yaitu *Linear Regression*, *Random Forest*, dan *XGBoost*. Ketiga algoritma tersebut digunakan karena berdasarkan penelitian sebelumnya, algoritma tersebut memiliki nilai akurasi yang tinggi pada setiap penelitiannya.

Penelitian ini bertujuan untuk mencari algoritma *Machine Learning* dengan nilai akurasi prediksi tertinggi untuk memprediksi nilai APS di Indonesia. Penelitian ini diharapkan dapat memberikan manfaat bagi pembuat kebijakan di sektor pendidikan dalam menetapkan kebijakan yang tepat untuk meningkatkan APS di Indonesia. Selain itu, penelitian ini juga diharapkan dapat meningkatkan pemahaman dan penerapan teknologi *Machine Learning* dalam analisis data pendidikan di Indonesia.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dijabarkan, maka dapat diidentifikasi beberapa masalah sebagai berikut:

1. Masih rendah dan belum meratanya Angka Partisipasi Sekolah (APS) di Indonesia, yang menunjukkan belum tercapainya pemerataan akses pendidikan di seluruh wilayah dan kelompok usia di Indonesia.
2. Belum banyak penelitian yang secara khusus memprediksi APS di Indonesia menggunakan algoritma *Machine Learning*.
3. Belum diketahui algoritma *Machine Learning* mana yang paling akurat untuk memprediksi APS.

1.3 Pembatasan Masalah

Penelitian mengenai analisis perbandingan algoritma *Machine Learning* untuk memprediksi APS memiliki beberapa batasan, yaitu:

1. Penelitian ini akan fokus pada algoritma *Machine Learning* (ML) *Linear Regression*, *Random Forest*, dan *XGBoost*;
2. Analisis akan dilakukan pada 34 provinsi di Indonesia;
3. Penelitian ini tidak menyertakan Provinsi Papua Barat Daya, Papua Tengah, Papua Selatan, dan Papua Pegunungan dalam analisis karena keempat provinsi tersebut baru resmi terbentuk pada tahun 2022, sehingga tidak

memiliki data historis yang lengkap dan konsisten sepanjang rentang waktu pengamatan (2003–2024);

4. Data yang digunakan dalam penelitian ini mencakup APS di 34 provinsi di Indonesia dari tahun 2003 – 2024;
5. Faktor-faktor yang dipertimbangkan dalam model prediksi terbatas pada data yang tersedia dari sumber resmi *website* Badan Pusat Statistik (BPS) dari tahun 2003 - 2024.

1.4 Perumusan Masalah

Berdasarkan latar belakang dan identifikasi masalah di atas, perumusan masalah penelitian ini adalah sebagai berikut: “Bagaimana perbandingan algoritma *Linear Regression*, *Random Forest*, dan *XGBoost* untuk memprediksi nilai APS di Indonesia?”

1.5 Tujuan Penelitian

Tujuan penelitian ini adalah untuk:

1. Mengidentifikasi algoritma *Machine Learning* dalam memprediksi APS di Indonesia;
2. Mencari algoritma *Machine Learning* dengan nilai akurasi prediksi tertinggi untuk memprediksi nilai APS di Indonesia.

1.6 Manfaat Penelitian

Manfaat penelitian ini adalah untuk:

1. Diharapkan dapat membantu kepada pembuat kebijakan di sektor pendidikan dalam menetapkan kebijakan yang tepat untuk meningkatkan APS di Indonesia;
2. Meningkatkan pemahaman dan penerapan teknologi *Machine Learning* dalam analisis data pendidikan di Indonesia.

Intelligentia - Dignitas