

## BAB I

# PENDAHULUAN

### 1.1 Latar Belakang

Pendidikan merupakan pilar utama dalam membentuk sumber daya manusia berkualitas. Di era digital, kemampuan berpikir analitis, logis, serta terampil dalam bertindak dan merespons isu sosial sangat penting dimiliki generasi penerus bangsa. Kemampuan ini dapat diukur melalui literasi membaca, numerasi, dan sains. Namun, hasil *Programme for International Student Assessment* (PISA) selama 22 tahun terakhir menunjukkan Indonesia belum mengalami peningkatan signifikan pada ketiga aspek tersebut. Data tahun 2022 mencatat skor literasi membaca sebesar 359, numerasi 366, dan literasi sains 383, yang masih jauh di bawah rata-rata negara OECD, yaitu 465-475 [1].

Untuk mengukur capaian belajar di tingkat daerah, Daerah Istimewa Yogyakarta menggunakan Asesmen Standardisasi Pendidikan Daerah (ASPD) sebagai evaluasi kompetensi literasi membaca, numerasi, dan sains siswa SMP/MTs. Nilai ASPD menjadi indikator penting untuk menilai mutu pendidikan serta dasar penyusunan kebijakan intervensi pembelajaran. Namun demikian, pemanfaatan data nilai ASPD untuk intervensi yang tepat waktu masih belum optimal, karena analisis yang mendalam terhadap nilai tersebut belum banyak dilakukan. Hal ini menyulitkan guru dan lembaga pendidikan dalam merancang strategi pembelajaran yang efektif.

Seiring dengan berkembangnya teknologi informasi dan data science, berbagai algoritma pembelajaran mesin mulai digunakan untuk membantu memprediksi hasil akademik siswa secara lebih akurat. Salah satu contohnya adalah penggunaan K-Nearest Neighbor (KNN) untuk memprediksi kelulusan ujian nasional di SMPN 2 Pagedangan[2], yang meskipun cukup akurat, memiliki keterbatasan dalam mengolah data dengan jumlah fitur yang banyak dan hubungan

antar variabel yang kompleks. Oleh karena itu, diperlukan pendekatan yang lebih adaptif dan mampu menangani kompleksitas data pendidikan dengan lebih baik.

Penelitian ini menggunakan algoritma *Decision Tree* dan *Random Forest* karena keduanya termasuk dalam *supervised learning* yang sangat cocok untuk data tabular seperti nilai-nilai akademik. *Decision Tree* memberikan hasil yang mudah diinterpretasi dalam bentuk pohon keputusan, sehingga memudahkan guru maupun pengambil kebijakan dalam memahami pola yang terbentuk dari data siswa. Sementara itu, *Random Forest* merupakan pengembangan dari *Decision Tree* dengan pendekatan ensemble, yang membuatnya lebih tahan terhadap *overfitting* dan menghasilkan akurasi yang lebih tinggi. Pada penelitian sebelumnya menunjukkan bahwa *Random Forest* mampu mencapai akurasi hingga 90.7% dalam memprediksi performa akademik siswa di Universitas Sebelas Maret[3]. Penelitian lain juga membuktikan bahwa algoritma ini unggul dalam menangani data yang kompleks dan tidak seimbang[4].

Selain pemilihan algoritma, penelitian ini juga mengintegrasikan metode optimasi *Simulated Annealing* (SA) untuk seleksi fitur. SA merupakan algoritma metaheuristic yang meniru proses pendinginan logam, dan dikenal efektif dalam mengeksplorasi ruang pencarian solusi secara luas tanpa mudah terjebak pada local optimum. Dengan menggunakan SA, proses seleksi fitur menjadi lebih efisien dan dapat menghasilkan kombinasi variabel input yang optimal[5], sehingga diharapkan dapat meningkatkan akurasi model dalam memprediksi nilai ASPD siswa. Pendekatan ini tidak hanya berfokus pada keakuratan prediksi, tetapi juga pada efisiensi model dan relevansi fitur yang digunakan dalam pengambilan keputusan pendidikan.

## 1.2 Rumusan Masalah

Penelitian ini berfokus pada evaluasi kinerja *Random Forest* dan *Decision Tree* dalam memprediksi nilai ASPD, serta efektivitas optimasi *Simulated Annealing*. Berdasarkan hal tersebut, rumusan masalah dalam penelitian ini adalah:

1. Model mana yang lebih unggul antara *Random Forest* dan *Decision Tree* dalam memprediksi nilai ASPD Siswa Menengah Pertama ?
2. Seberapa efektif metode optimasi *Simulated Annealing* dalam meningkatkan kinerja *Random Forest* dan *Decision Tree*?

### 1.3 Batasan Masalah

Dalam rangka memperoleh gambaran yang jelas mengenai masalah yang dibahas, perlu mempertimbangkan keterbatasan seperti kemampuan, waktu, dan biaya. Berikut batasan-batasan yang diberikan dalam penelitian ini:

#### 1. Lokasi Penelitian

Penelitian ini dilakukan dengan menggunakan dataset nilai tryout dan nilai ASPD siswa dari SMP Negeri 2 Mlati.

#### 2. Data yang Digunakan

- a. Variabel bebas (fitur) yang digunakan dalam penelitian ini meliputi: 'bahasa indonesia2', 'ASPD Literasi', 'rata-rata Literasil', 'ASPD Sains', 'ASPD Numerasi', 'nilai TO1'.

- b. Variabel terikat (target prediksi) adalah nilai ASPD siswa.

#### 3. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan penerapan model prediksi *Random Forest* dan *Decision Tree*.

#### 4. Optimasi Parameter

Proses optimasi model dilakukan menggunakan metode *Simulated Annealing* untuk meningkatkan performa prediksi pada kedua model tersebut.

#### 5. Asumsi Penelitian

Dataset telah melalui proses pembersihan, normalisasi, dan penanganan *outlier*. Nilai tryout dan ASPD dianggap mencerminkan kemampuan akademik siswa. Faktor eksternal seperti kondisi psikologis atau sosial siswa diasumsikan tidak berpengaruh signifikan terhadap hasil ASPD.

#### 6. Batasan Data dan Generalisasi

Penelitian ini hanya berfokus pada variabel akademik yang disebutkan tanpa mempertimbangkan faktor lain seperti pelajaran di luar variabel yang dianalisis.

#### 7. Batasan Temporal

Data yang dianalisis berasal dari periode tertentu yaitu data tahun 2024 sesuai dataset yang tersedia.

### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, tujuan penelitian ini adalah:

1. Menilai dan membandingkan kinerja dua model prediksi, yaitu *Random Forest* dan *Decision Tree*, dalam memprediksi nilai ASPD siswa SMP Negeri 2 Mlati berdasarkan data nilai sebelumnya.
2. Menilai seberapa efektif metode optimasi *Simulated Annealing* dalam meningkatkan performa dan akurasi model prediksi *Random Forest* dan *Decision Tree*.

### 1.5 Manfaat Penelitian

1. Memberikan kontribusi pada pengembangan metode prediksi berbasis *Random Forest* dan *Decision Tree* serta penerapan optimasi *Simulated Annealing* dalam konteks pendidikan.
2. Memperdalam pemahaman mengenai peran teknik optimasi dalam meningkatkan performa model prediksi di bidang pendidikan.

### 1.6 Sistematika Penulisan

Berikut adalah sistematika penulisan skripsi yang lebih singkat:

## BAB I PENDAHULUAN

Berisi latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta metode penelitian yang digunakan.

**BAB II TINJAUAN PUSTAKA**

Memuat tinjauan penelitian sebelumnya, dasar teori yang digunakan, serta kerangka pemikiran yang mendasari penelitian ini.

**BAB III METODE PENELITIAN**

Menjelaskan desain penelitian, objek penelitian, metode pengumpulan data, teknik analisis data, dan instrumen yang digunakan.

**BAB IV HASIL DAN PEMBAHASAN**

Berisi hasil penelitian yang diperoleh dan analisis serta pembahasan atas temuan-temuan yang ada.

**BAB V PENUTUP**

Menyajikan kesimpulan dari penelitian, saran untuk penelitian selanjutnya, dan keterbatasan yang ada dalam penelitian ini.