

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Perkembangan teknologi saat ini telah membuka berbagai peluang bagi beragam sektor, termasuk sektor pendidikan. Kemajuan di bidang teknologi informasi telah memudahkan pengolahan dan analisis data dalam skala besar melalui teknik seperti data mining. Data mining telah banyak dimanfaatkan dalam berbagai bidang untuk menggali informasi tersembunyi yang ada di dalam data, seperti memprediksi tren, memahami pola, dan mengambil keputusan berbasis data (Harahap & Sulindawaty, 2020). Khusus dalam sektor pendidikan, penerapan data mining berpotensi memberikan wawasan baru mengenai dinamika pembelajaran serta pemahaman yang lebih mendalam terkait prestasi siswa. Dengan informasi yang lebih akurat, sekolah dapat meningkatkan kualitas pendidikan, memantau kemajuan belajar siswa, dan memberikan dukungan yang lebih tepat sasaran kepada para siswa (Wiratama & Pradnya, 2022).

Dalam konteks pendidikan, pemanfaatan data mining mulai diterapkan dalam pemantauan prestasi belajar siswa. Teknik ini mampu mengidentifikasi berbagai faktor yang memengaruhi prestasi akademik dan membantu pihak sekolah untuk menyusun kebijakan yang tepat. Prediksi prestasi belajar siswa, misalnya, dapat dilakukan dengan algoritma seperti C4.5 yang terbukti efektif dalam menghasilkan model prediksi dengan tingkat akurasi yang tinggi. Algoritma ini memanfaatkan proses pembentukan pohon keputusan yang sederhana namun mampu memberikan hasil yang relevan. Prediksi yang dihasilkan dapat memberikan gambaran awal kepada pihak sekolah mengenai siswa-siswa yang memerlukan perhatian khusus, sehingga intervensi yang dilakukan lebih tepat sasaran.

Di SMK Bina Warga 2 Kota Bogor, prestasi siswa menjadi salah satu fokus utama karena tingginya harapan masyarakat terhadap lulusan yang mampu bersaing di dunia kerja. Untuk mencapai hal tersebut, pihak sekolah berusaha melakukan pemantauan yang lebih mendalam terhadap faktor-faktor yang memengaruhi prestasi siswa. Namun, proses ini sering

kali menghadapi kendala karena keterbatasan metode kurang optimal yang kurang akurat dan memakan waktu. Hal ini mengakibatkan kesulitan dalam mengelola data siswa yang terus bertambah dari tahun ke tahun. Sebagai contoh, faktor-faktor seperti tingkat kehadiran, partisipasi dalam kegiatan ekstrakurikuler, dan tingkat dukungan dari lingkungan keluarga sering kali terabaikan dalam analisis kurang optimal.

Pengelolaan data secara kurang optimal memiliki sejumlah tantangan yang signifikan. Salah satunya adalah waktu yang diperlukan untuk mengumpulkan, menyortir, dan menganalisis data dalam jumlah besar. Ketergantungan pada metode kurang optimal juga meningkatkan risiko kesalahan manusia, seperti kelalaian dalam pencatatan atau interpretasi data. Selain itu, sulitnya mengintegrasikan data dari berbagai sumber, seperti catatan kehadiran, nilai ujian, dan laporan kegiatan ekstrakurikuler, seringkali menghambat pihak sekolah dalam mendapatkan gambaran yang utuh tentang performa siswa. Tantangan-tantangan ini menunjukkan perlunya solusi berbasis teknologi yang dapat mengotomatisasi proses pengelolaan data secara efisien.

Algoritma C4.5 adalah salah satu metode dalam data mining yang digunakan untuk membangun model prediksi berbasis pohon keputusan. Algoritma ini bekerja dengan membagi data menjadi cabang-cabang berdasarkan atribut yang paling signifikan, menggunakan Gain Ratio sebagai kriteria pemilihan atribut terbaik. Proses ini dimulai dengan membangun akar pohon yang merepresentasikan atribut utama, diikuti dengan pembentukan cabang-cabang yang mewakili nilai-nilai atribut tersebut. Pada setiap langkah, algoritma terus membagi data hingga mencapai kondisi di mana data dapat dikelompokkan secara jelas ke dalam kategori tertentu. Keunggulan utama algoritma C4.5 adalah kemampuannya untuk menangani data dengan atribut kontinu dan diskrit, serta kemampuannya untuk menangani data yang mengandung nilai yang hilang (missing values). Dengan pendekatan ini, algoritma C4.5 dapat menghasilkan model prediksi yang tidak hanya akurat, tetapi juga mudah dipahami oleh pengguna non-teknis, seperti guru dan staf administrasi.

SMK Bina Warga 2 Kota Bogor sebagai salah satu lembaga pendidikan vokasional terkemuka di Kota Bogor memiliki sejumlah data penting terkait latar belakang, aktivitas,

dan hasil belajar siswa. Kendala yang sering dihadapi pihak sekolah adalah sulitnya mengidentifikasi dan memetakan siswa yang berpotensi mengalami kesulitan dalam pencapaian akademik. Selain itu, banyaknya faktor yang memengaruhi prestasi belajar seperti kehadiran, interaksi siswa, dan dukungan lingkungan belajar, menjadi tantangan tersendiri bagi pihak sekolah untuk mengelola data yang kompleks. Tanpa adanya sistem pendukung yang kuat, pihak sekolah akan kesulitan melakukan pemantauan dan intervensi secara tepat waktu. Tantangan ini memunculkan kebutuhan akan pendekatan yang lebih terstruktur dan berbasis teknologi untuk mendukung pemantauan prestasi belajar siswa.

Penerapan data mining melalui algoritma C4.5 diharapkan menjadi solusi efektif untuk mengoptimalkan prediksi prestasi belajar siswa di SMK Bina Warga 2 Kota Bogor. Algoritma ini memungkinkan sekolah mengelompokkan dan menganalisis data siswa secara mendalam berdasarkan variabel-variabel relevan, sehingga model prediksi dapat dibangun untuk mengidentifikasi siswa yang membutuhkan perhatian khusus (Setio et al., 2020). Algoritma C4.5 juga memiliki keunggulan dalam menangani data yang memiliki nilai atribut kontinu maupun diskrit, sehingga fleksibel untuk diaplikasikan pada berbagai jenis data siswa. Dengan optimasi algoritma C4.5, pihak sekolah dapat melakukan intervensi dini dan meningkatkan layanan pendidikan untuk memenuhi harapan masyarakat terhadap lulusan yang kompeten.

Masalah utama dalam penelitian ini adalah kurangnya kejelasan mengenai karakteristik dan variabel utama yang memengaruhi prestasi belajar siswa, seperti tingkat kehadiran, hasil tugas, dan latar belakang keluarga, sehingga analisis yang dilakukan belum akurat. Selain itu, belum adanya analisis prediktif untuk memetakan siswa yang berpotensi mengalami kesulitan belajar menyebabkan intervensi yang dilakukan sering kali terlambat dan kurang efektif. Kurangnya kajian untuk mengurangi peserta didik yang tidak relevan, terutama dalam konteks keilmuan sistem informasi, juga menjadi kendala karena tidak diketahui secara jelas variabel-variabel yang berkontribusi terhadap relevansi data siswa. Selain itu, absennya penerapan algoritma seperti C4.5 dalam pengolahan data menyebabkan prediksi prestasi belajar siswa belum mencapai tingkat akurasi yang optimal.

Selain itu, penerapan algoritma C4.5 diharapkan mampu memberikan rekomendasi yang lebih terarah bagi guru dan wali kelas dalam mendampingi siswa. Melalui analisis data yang komprehensif, pihak sekolah dapat mengidentifikasi pola-pola tertentu yang berkontribusi pada keberhasilan atau kegagalan siswa dalam mencapai target pembelajaran. Misalnya, siswa yang sering absen dan memiliki nilai rendah pada mata pelajaran tertentu dapat segera mendapatkan perhatian khusus, seperti bimbingan belajar tambahan atau sesi konsultasi dengan guru.

Melalui penelitian berjudul **“Analisis dan Optimasi Algoritma C4.5 Untuk Memprediksi Prestasi Belajar Siswa di SMK Bina Warga 2 Bogor”** diharapkan solusi praktis ini dapat diterapkan untuk mendukung keberhasilan akademik siswa. Penelitian ini juga bertujuan untuk memberikan kontribusi nyata dalam pengembangan sistem pendidikan berbasis data, yang tidak hanya relevan bagi SMK Bina Warga 2 Kota Bogor, tetapi juga dapat menjadi model bagi institusi pendidikan lainnya.

## **1.2. Identifikasi Masalah**

Berdasarkan latar belakang yang telah diuraikan sebelumnya, dapat diidentifikasi beberapa permasalahan dalam pembuatan tugas akhir ini, antara lain:

1. Tidak diketahui dengan jelas karakteristik dan variabel-variabel utama yang memengaruhi prestasi belajar siswa, seperti tingkat kehadiran, hasil tugas, dan latar belakang keluarga, sehingga sulit untuk melakukan analisis secara akurat.
2. Belum adanya analisa prediksi untuk memetakan siswa dengan potensi kesulitan belajar, sehingga upaya intervensi cenderung terlambat dan kurang efektif.
3. Belum adanya kajian dan upaya untuk mengurangi peserta didik tidak relevan terutama dari segi keilmuan sistem informasi serta tidak diketahui dengan jelas karakteristik dan variabel-variabel yang dapat mempengaruhi data siswa bisa menjadi non relevan.
4. Belum adanya algoritma seperti C4.5 untuk pengolahan data, sehingga prediksi prestasi belajar belum mencapai tingkat akurasi yang diharapkan.

### 1.3. Rumusan Masalah

Dalam upaya mengoptimalkan prediksi prestasi belajar siswa di SMK Bina Warga 2 Kota Bogor, diperlukan pemahaman yang mendalam terhadap permasalahan yang dihadapi dalam proses pemantauan dan prediksi. Berdasarkan latar belakang yang ada, berikut ini adalah rumusan masalah yang akan dijawab dalam penelitian ini:

1. Apa saja karakteristik dan variabel-variabel utama yang memengaruhi prestasi belajar siswa di SMK Bina Warga 2 Kota Bogor?
2. Bagaimana merancang sistem prediksi yang dapat membantu mengidentifikasi siswa dengan potensi kesulitan belajar sejak dini?
3. Bagaimana mengelola data prestasi siswa yang kompleks agar dapat dimanfaatkan secara optimal untuk mendukung pengambilan keputusan?
4. Bagaimana merancang algoritma C4.5 untuk meningkatkan akurasi prediksi prestasi belajar siswa?

### 1.4. Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengembangkan dan mengoptimalkan model prediksi prestasi belajar siswa di SMK Bina Warga 2 Kota Bogor menggunakan algoritma C4.5. Berdasarkan rumusan masalah yang telah diidentifikasi, tujuan penelitian ini adalah:

1. Mengidentifikasi karakteristik dan variabel-variabel utama yang memengaruhi prestasi belajar siswa di SMK Bina Warga 2 Kota Bogor.
2. Menganalisis cara untuk mengidentifikasi siswa yang memiliki potensi kesulitan belajar berdasarkan data prestasi yang tersedia.
3. Mengolah data prestasi siswa yang kompleks untuk menghasilkan informasi yang mendukung analisis prestasi belajar.
4. Mengoptimalkan penggunaan algoritma C4.5 untuk meningkatkan akurasi prediksi prestasi belajar siswa.

### 1.5. Batasan Masalah

Batasan masalah dalam penelitian ini difokuskan pada penerapan algoritma C4.5 untuk prediksi prestasi belajar siswa di SMK Bina Warga 2 Kota Bogor, dengan mempertimbangkan faktor-faktor yang memengaruhi prestasi akademik siswa. Adapun ruang lingkup penelitian ini dibatasi pada data yang tersedia di sekolah serta variabel yang relevan untuk model prediksi:

1. Penelitian ini tidak akan merancang aplikasi untuk melakukan prediksi prestasi siswa.
2. Penelitian ini hanya akan melakukan analisis prediksi menggunakan algoritma C4.5 tanpa membandingkannya dengan algoritma lainnya.
3. Sumber data yang akan digunakan dalam penelitian ini adalah data siswa SMK Bina Warga 2 Kota Bogor, yang mencakup data nilai akademik, kehadiran, dan faktor-faktor lain yang relevan terbatas periode Tahun Ajaran 2024-2025.

### 1.6. Manfaat Penelitian

Manfaat penelitian ini dibagi menjadi tiga, yaitu bagi objek penelitian (SMK Bina Warga 2 Kota Bogor), bagi pengembangan ilmu pengetahuan, dan bagi kampus. Adapun manfaat penelitian ini adalah sebagai berikut:

**a. Bagi Objek Penelitian (SMK Bina Warga 2 Kota Bogor)**

Penelitian ini diharapkan dapat memberikan solusi praktis dalam meningkatkan akurasi prediksi prestasi belajar siswa. Dengan menggunakan algoritma C4.5, pihak sekolah dapat melakukan pemantauan yang lebih efektif dan mengidentifikasi siswa yang membutuhkan perhatian khusus, sehingga strategi pembelajaran dapat disesuaikan untuk mendukung pencapaian akademik yang lebih baik.

**b. Bagi Pengembangan Ilmu Pengetahuan**

Penelitian ini memberikan kontribusi dalam mengembangkan penerapan data mining, khususnya algoritma C4.5, dalam konteks pendidikan. Temuan dari penelitian ini diharapkan dapat memperkaya kajian mengenai pemanfaatan

teknologi untuk prediksi prestasi belajar siswa dan membuka peluang penelitian lebih lanjut di bidang serupa.

**c. Bagi Kampus**

Penelitian ini juga memberikan manfaat bagi kampus dalam hal penerapan penelitian berbasis teknologi dan data mining di dunia pendidikan. Hasil penelitian dapat dijadikan referensi untuk pengembangan kurikulum atau proyek penelitian lanjutan yang relevan, serta meningkatkan reputasi kampus dalam bidang penelitian terapan yang berdampak langsung pada sektor pendidikan.

### **1.7. Sistematika Penulisan**

Sistematika penulisan penelitian ini disusun untuk memberikan Gambaran umum tentang penelitian yang akan dijalankan, sistematika penulisan ini adalah sebagai berikut:

#### **BAB I PENDAHULUAN**

Bab ini memuat secara garis besar tentang mengenai latar belakang, perumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

#### **BAB II TINJAUAN PUSTAKA**

Bab ini akan menjelaskan uraian mengenai tinjauan umum yang berkaitan dengan objek yang sedang diteliti, landasan teori yang berhubungan dengan penelitian yang akan dilakukan, memuat dari hasil penelitian sejenis terdahulu yang menginspirasi atau melandasi pelaksanaan penelitian

#### **BAB III METODOLOGI PENELITIAN**

Pada bab ini dijelaskan tentang metode penelitian yang digunakan dalam proses pengumpulan dan analisis data. Metode penelitian ini disusun untuk mendukung proses penelitian sehingga mencapai tujuan yang diharapkan.

#### **BAB IV ANALISIS DAN PEMBAHASAN**

Pada bab ini, dilakukan analisis terhadap data yang telah dikumpulkan serta pembahasan hasil penelitian berdasarkan metode yang telah diterapkan. Proses analisis bertujuan untuk menginterpretasikan data secara sistematis guna memperoleh temuan yang relevan dengan tujuan penelitian. Selain itu, dalam bab ini juga dibahas penerapan metode penelitian yang digunakan dalam pengolahan data, termasuk teknik klasifikasi dengan algoritma C4.5, serta evaluasi hasil yang diperoleh untuk memastikan keakuratan model prediksi yang dihasilkan.

#### **BAB V PENUTUP**

Penelitian ini menyimpulkan bahwa penerapan algoritma C4.5 dalam prediksi prestasi belajar siswa di SMK Bina Warga 2 Kota Bogor mampu mengidentifikasi faktor-faktor yang berpengaruh dengan tingkat akurasi yang cukup tinggi, sehingga dapat membantu pihak sekolah dalam melakukan intervensi secara lebih tepat. Namun, untuk meningkatkan efektivitas model, perlu dilakukan pengembangan lebih lanjut dengan mempertimbangkan variabel tambahan serta integrasi sistem prediksi ke dalam sistem informasi sekolah agar lebih praktis dan bermanfaat dalam pengambilan keputusan akademik.