

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

*Knowledge Discovery in Databases* (KDD) adalah penerapan metode penelitian pada Data Mining. Knowledge Discovery in Databases (KDD) yaitu tahapan yang dilakukan dalam menemukan pengetahuan dari sekumpulan data. Proses pencarian informasi tersembunyi dalam suatu basis data yang besar sering menggunakan istilah Data Mining dan *knowledge discovery in databases* (KDD) (Simanullang, *et al.* 2022).

Data Mining adalah teknik analisis untuk menemukan korelasi, pola, tren dari sejumlah besar data itu disimpan dalam *database*. Klasifikasi, pengelompokan, dan asosiasi adalah Algoritma Data Mining yang paling banyak digunakan. Data Mining dapat dipertimbangkan lebih fleksibel bila dibandingkan dengan matematika yang ketat pendekatan metode statistik tradisional (Gezer, *et al.* 2022). Teknik yang mendukung dan paling banyak digunakan dalam Data Mining adalah Machine Learning (Aprihartha, *et al.* 2024). Machine Learning merupakan bagian terpenting dari bidang Artificial Intelligence (AI), berkembang pesat dan secara bertahap merambah ke berbagai industri (Ping Li, *et al.* 2024).

Algoritma C4.5 merupakan salah satu Algoritma yang digunakan untuk membentuk pohon keputusan berdasarkan data latih (Pramudito, 2022). Pohon keputusan sering digunakan untuk mengatasi masalah klasifikasi di bidang Data Mining dan Machine Learning, karena banyak manfaatnya, termasuk arsitekturnya yang jelas dan sederhana, kualitas yang sangat baik, dan ketahanan (Aaboub, 2023). Dalam proses pembuatan pohon keputusan, mencari variabel dengan informasi maksimum dalam kumpulan data untuk membentuk *node* data, dan kemudian menetapkan cabang sesuai dengan karakteristik variabel, dan setiap *subset* cabang mengulangi proses pembentukan cabang dan *node* hingga selesainya seluruh pohon

keputusan, dan setiap jalur pohon keputusan mewakili aturan klasifikasi, dan data baru diprediksi (Fengwei Lu, *et al.* 2023). Algoritma C4.5 dikenal dengan akurasi dan kecepatan klasifikasinya yang tinggi, memanfaatkan perolehan informasi rasio untuk memilih atribut. Selain itu, ia dapat menangani atribut *diskrit* dan *kontinu*, serta data yang tidak lengkap (Wang, *et al.* 2024). Algoritma C4.5 menggunakan *information gain rate* sebagai atribut untuk membagi *subset* (Ping Li, *et al.* 2024).

Pendidikan merupakan kegiatan yang sangat kompleks. Dalam kehidupan, pendidikan berperan penting untuk menjamin kelangsungan hidup manusia, karena pendidikan merupakan sebuah fasilitas untuk mengembangkan dan meningkatkan kualitas sumber daya manusi (Marita dan Prayogi, 2024). Salah satu faktor terbesar yang menjadi kendala dalam dunia pendidikan adalah keterbatasan biaya yang membuat banyak keluarga dari ekonomi lemah tidak melanjutkan pendidikan ke jenjang yang lebih tinggi (Husain, *et al.* 2023).

Beasiswa adalah sebuah program dalam wujud bantuan keuangan yang ditujukan pada individu untuk melanjutkan pendidikan mereka dengan tujuan membantu mengurangi beban keuangan selama masa studi, terutama dalam situasi sulit, sehingga dapat membantu memperlancar proses belajar (Surya Maulana, *et al.* 2023). Beasiswa bertujuan membantu siswa yang berbakat dan berprestasi dari kalangan ekonomi kurang mampu agar dapat melanjutkan pendidikan ke jenjang yang lebih tinggi. Terdapat berbagai jenis beasiswa, dan salah satu beasiswa yang saat ini diberikan oleh pemerintah Negara Indonesia adalah Program Indonesia Pintar (Kadja, *et al.* 2023).

Program Indonesia Pintar adalah salah satu langkah yang dilakukan pemerintah Indonesia pada era kepemimpinan Presiden Ir. Joko Widodo, sesuai dengan Peraturan Presiden No. 7 Tahun 2014. Salah satu adalah pemberian Kartu Indonesia Pintar, yang bertujuan memberikan bantuan kepada siswa kurang mampu secara finansial agar dapat memperoleh pendidikan berkualitas. Program ini diharapkan dapat mencegah putus sekolah dikalangan anak-anak dari keluarga miskin dan memenuhi kebutuhan sekolah siswa, seperti biaya SPP, uang saku, dan buku pelajaran. Dengan demikian, siswa miskin dapat memiliki kesempatan yang sama dengan siswa lainnya untuk mengakses pendidikan. Diharapkan, program ini dapat menciptakan masyarakat yang lebih berdaya saing, cerdas, dan unggul di bidang pendidikan (Maajid Amadi, 2023).

Program Indonesia Pintar merupakan inisiatif pemerintah yang memberikan bantuan keuangan kepada siswa-siswa yang memenuhi kriteria tertentu, sehingga mereka dapat melanjutkan pendidikan tanpa hambatan ekonomi. Meskipun Program Indonesia Pintar memiliki dampak positif yang signifikan, tantangan dalam pelaksanaannya terletak pada proses seleksi penerima bantuan (Fazira, *et al.* 2023). Syarat-syarat khusus yang harus dipenuhi calon penerima beasiswa tersebut meliputi pekerjaan orang tua, penghasilan orang tua, status penerima KPS, dan status penerima KIP. Seluruh variabel ini harus dihitung agar penetapan penerima beasiswa sesuai dengan yang dipersyaratkan (Yanasari dan Arifin, 2023).

Dinas Pendidikan Kabupaten Solok Selatan, sebagai salah satu SKPD di Kabupaten Solok Selatan, bertanggung jawab atas jalannya tujuan serta fungsi pendidikan dalam wilayah kerjanya. Dinas ini memiliki peran dan tujuan untuk membina serta menciptakan pendidikan yang baik, sebagaimana tertulis dalam visi dan misinya. Dinas Pendidikan Kabupaten Solok Selatan juga bertugas mengelola data pendidikan di wilayah tersebut, termasuk pengelolaan data untuk pemberian beasiswa Program Indonesia Pintar. Banyaknya peserta didik menyebabkan penentuan calon penerima beasiswa dengan proses seleksi manual dirasa kurang efektif. Oleh karena itu, diperlukan prediksi untuk menentukan penerima beasiswa dengan menerapkan metode Algoritma C4.5, yang akan membantu menyeleksi siswa yang berhak menerima beasiswa Program Indonesia Pintar. Seluruh data akan diuji dan diolah menggunakan aplikasi *software* RapidMiner dalam proses pengujian Data Mining dan penerapan Algoritma C4.5.

Nilai prediksi tersebut dihitung melalui klasifikasi Algoritma C4.5 menggunakan model pohon keputusan yang dapat mencari variabel dengan informasi maksimum dalam kumpulan data untuk membentuk *node* data, dan kemudian menetapkan cabang sesuai dengan karakteristik variabel, dan setiap *subset* cabang mengulangi proses pembentukan cabang dan *node* hingga selesainya seluruh pohon keputusan, dan setiap jalur pohon keputusan mewakili aturan klasifikasi, dan data baru diprediksi oleh aturan klasifikasi (Fengwei Lu, *et al.* 2023).

Menurut penelitian terdahulu terkait dengan Penerapan Algoritma C4.5 Mengklarifikasi Penerimaan Bantuan Sosial Menggunakan *Feature Selection* dalam pemilihan bantuan KIP didapatkannya nilai akurasi dari pemodelan pohon keputusan

memakai Algoritma C4.5 atau *decision tree* dengan tingkat keberhasilan Penerimaan KIP. Evaluasi ini memakai *confusion matrix* untuk diperoleh nilai akurasi sebesar 98.21%, presisi sebesar 98.21%, *recall* sebesar 99.48% (Wirawan, *et al* 2023).

Penelitian lainnya dengan studi kasus di SMK Plus Al-Hilal Arjawinangun. Penelitian ini membahas tentang Penerapan Algoritma C4.5 Untuk Seleksi Penjurusan Siswa Baru Pada Sekolah Menengah Kejuruan. Hasil yang di dapat memberikan manfaat untuk sekolah menjadi lebih cepat dalam pengklasifikasian penentuan jurusan siswa SMK Plus Al-Hilal Arjawinangun. Dari pemodelan klasifikasi menggunakan Algoritma C4.5 tingkat akurasi mencapai 98.02%, presisi 98.73% dan *recall* 98.73% (Roghib, *et al.* 2024).

Penelitian lainnya membahas tentang Konstruksi dan Optimalisasi Sistem Pendukung Keputusan Mesin Penjual Otomatis berdasarkan pohon keputusan C4.5 yang ditingkatkan. Hasil penelitian menunjukkan bahwa Akurasi Algoritma C4.5 dapat mencapai tingkat tertinggi 87% dan terendah sebesar 68%. Akurasi Algoritma C4.5 yang ditingkatkan dapat mencapai 87% pada tingkat tertinggi dan 67% pada tingkat terendah, dengan sedikit perbedaan di antara keduanya. Waktu berjalan maksimum dari Algoritma C4.5 yang ditingkatkan adalah sekitar 5500 ms, dan waktu minimum mendekati 1 ms. Selain itu, waktu berjalan ketujuh dataset lebih baik dibandingkan dengan Algoritma yang tidak dimodifikasi (Ping Li, *et al.* 2024).

Berdasarkan permasalahan dan penelitian terdahulu yang telah diuraikan di atas, penulis tertarik untuk menjadikan permasalahan tersebut sebagai fokus penelitian dan menyusunnya dalam bentuk tesis dengan judul “Implementasi Algoritma C4.5 Untuk Prediksi Penerima Beasiswa Program Indonesia Pintar (Studi Kasus di Wilayah Dinas Pendidikan Kabupaten Solok Selatan)”.

## **1.2 Perumusan Masalah**

Berdasarkan uraian latar belakang masalah yang dijelaskan di atas maka dapat dirumuskan pokok permasalahan dalam penelitian ini sebagai berikut:

1. Bagaimana penerapan Algoritma C4.5 dapat memprediksi penerima Beasiswa Program Indonesia Pintar?

2. Bagaimana menguji aplikasi menggunakan Algoritma C4.5 dapat memprediksi penerima Beasiswa Program Indonesia Pintar?
3. Bagaimana menerapkan *software tools* dengan metode Algoritma C4.5 dapat memprediksi penerima Beasiswa Program Indonesia Pintar?

### 1.3 Batasan Masalah

Berdasarkan perumusan masalah di atas, untuk menghindari terlalu luasnya permasalahan dan pemecahan masalah maka perlu adanya batasan ruang lingkup agar penelitian menjadi lebih terarah dan sistematis. Batasan-batasan yang diberikan yaitu:

1. Penelitian ini menggunakan Algoritma C4.5 dalam proses penelitian dengan siswa sebagai objek.
2. Penelitian ini dilakukan di wilayah Kabupaten Solok Selatan dengan pengambilan data pada Sekolah Dasar melalui Dapodik. Sumber data diambil dari *database* online yang tersedia pada *website* Dinas Pendidikan Kabupaten Solok Selatan.
3. *Tools* yang digunakan dalam proses pengolahan data pada penelitian ini adalah *software* RapidMiner.

### 1.4 Tujuan Penelitian

Berdasarkan perumusan masalah dan batasan masalah di atas, maka tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Penerapan Algoritma C4.5 dapat memprediksi penerima beasiswa Program Indonesia Pintar.
2. Menguji aplikasi menggunakan Algoritma C4.5 dapat memprediksi penerima Beasiswa Program Indonesia Pintar.
3. Menerapkan *software tools* dengan metode Algoritma C4.5 dapat memprediksi penerima beasiswa Program Indonesia Pintar.

## **1.5 Manfaat Penelitian**

Selain bernilai tujuan, tentunya penelitian ini diharapkan dapat bernilai manfaat bagi beberapa pihak, berikut manfaat dari penelitian yang ingin dicapai:

1. Bagi Peneliti dapat menerapkan ilmu pengetahuan yang diperoleh tentang Algoritma C4.5 dan mengimplementasikannya di kehidupan nyata.
2. Bagi peneliti berikutnya dapat memanfaatkan penelitian ini untuk mengembangkan model atau Algoritma yang lebih optimal.
3. Bagi objek yang diteliti dapat memberikan manfaat kepada berbagai pihak yang terkait untuk memprediksi penerima beasiswa Program Indonesia Pintar di wilayah Dinas Pendidikan Kabupaten Solok Selatan.

## **1.6 Sistematika Penulisan**

Sub bab ini membahas sistematika penulisan yang merupakan gambaran umum setiap bab dalam tesis. Adapun sistematika penulisan tersebut adalah sebagai berikut.

## **BAB I PENDAHULUAN**

Pada BAB I ini berisikan mengenai Latar Belakang Penelitian, Perumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian dan Sistematika Penulisan.

## **BAB II LANDASAN TEORI**

Pada BAB II ini membahas tentang teori-teori dan penerapan metode yang berhubungan dengan penelitian tesis ini. Teori dan penerapan tersebut akan menjadi landasan dalam penyusunan tesis, seperti teori tentang penerapan Metode Algoritma C4.5 dan lain-lain.

## **BAB III METODOLOGI PENELITIAN**

Pada BAB III ini membahas tentang kerangka kerja penelitian yang menguraikan tahap-tahap yang dilakukan dalam penelitian.

#### **BAB IV ANALISA DAN PERANCANGAN**

Pada BAB IV ini membahas tentang bagaimana dalam menganalisa permasalahan dan menganalisa data terkait dengan metode Algoritma C4.5.

#### **BAB V IMPLEMENTASI DAN HASIL**

Pada BAB V ini membahas tentang pengujian data dan pengimplementasian dengan menggunakan *software* RapidMiner.

#### **BAB VI KESIMPULAN DAN SARAN**

Pada BAB VI ini membahas tentang kesimpulan dari hasil penelitian prediksi penerima beasiswa Program Indonesia Pintar yang diimplementasikan serta memberikan saran untuk penelitian selanjutnya.