

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

*Data mining* merupakan proses dalam mendapatkan informasi berguna yang berasal dari basis data yang besar. Dengan bantuan data mining, pengguna dapat mengekstraksi informasi tersembunyi yang berharga yang dapat membantu dalam pengambilan keputusan atau digunakan untuk membuat prediksi (Suweleh dkk, 2020). Dalam data mining, data ini akan diolah dengan beberapa teknik sehingga akan lebih mudah untuk dimengerti oleh orang lain. Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu: Deskripsi, Estimasi, Prediksi, material data mining sebenarnya sudah terbentuk karena faktor rutinitas dan waktu seraya perusahaan melakukan aktivitasnya (Simorangkir, 2019).

Penggunaan *data mining* dapat diimplementasikan di berbagai bidang, diantaranya di bidang bisnis, kesehatan, telekomunikasi, dan bidang pendidikan. Keluaran dari *data mining* bisa dipakai dalam pengambilan keputusan dimasa depan (Ependi & Akbar, 2021). Istilah lain yang sering dikemukakan untuk teknik ini adalah *Knowledge Discovery Database (KDD)*. Data yang diperoleh sering sekali bersifat abstrak atau sulit dimengerti serta terkadang memiliki dimensi yang sangat besar sehingga sangatlah sulit untuk mengetahui pesan atau makna dari data tersebut. Banyak metode data mining yang bisa digunakan untuk klasifikasi. Beberapa jenis algoritma dalam klasifikasi data mining yang digunakan antaranya adalah C4.5, ID3, *k-nearestneighbor classifier*, *Naïve Mayes*, *SVM*, dan *ANN* (Sugiyanto & Pujiyanita, 2019).

Saat suatu organisasi baik itu perusahaan maupun suatu institusi yang mempunyai data yang kompleks, tidak menutup kemungkinan banyak sekali informasi yang dapat diperoleh, serta bagaimana solusi data mining bisa diterapkan dengan berbagai teknik diantaranya yaitu *classification*, *association* dan *clustering* (Suhartini

& Yuliani, 2021). Algoritma C4.5 adalah algoritma data mining yang digunakan untuk melakukan klasifikasi dan estimasi. Algoritma C4.5 termasuk dalam Algoritma induksi decision tree yang dikembangkan oleh J. Ross Quinlan yang dikenal dengan nama ID3 (Suweleh Dkk, 2020). Pada sebuah penelitian seleksi penerimaan mahasiswa baru bahwa Algoritma C4.5 lebih konsisten dan lebih handal saat menguji data dibandingkan *Naïve Mayes* (Alejandrino dkk, 2020).

Sugianto, et al. (2019) dalam penelitian tentang prediksi kelulusan siswa dengan data sampel sebanyak 901 siswa yang sebelumnya 19 *field* dengan menghilangkan *field* no. Sehingga *field* yang digunakan menjadi 18 *field*. Dari penelitian ini terdapat 11 indikator yang mempengaruhi prediksi kelulusan SMKN 2 cimahi yaitu: Asal sekolah, Tahun Pelajaran, Nilai Ujian Nasional Bahasa Indonesia, Nilai Ujian Nasional Bahasa Inggris, Nilai Ujian Sekolah Bahasa Inggris, Nilai Ujian kompetensi, Nilai Ebtanas Murni SMP Bahasa Indonesia, Nilai Ebtanas Murni SMP Bahasa Inggris, Nilai Ebtanas Murni SMP Matematika, Nilai Ebtanas Murni SMP IPA, Jumlah.

Puspitorini & Adhi (2020) pada penelitian penerapan data mining dengan menggunakan algoritma C4.5 untuk analisa kepangkatan aparatur sipil negara menggunakan atribut: pendidikan terakhir, masa kerja, sasaran kerja pegawai (SKP), DP3, pangkat sebelumnya. hasil penelitian dengan model decision tree menggunakan algoritma C4.5 yang dievaluasi dengan confusion matix menghasilkan angka akurasi 80,00%, precision 83,33%, dan recall 80,00%. Disimpulkan hasil penelitian ini dapat memberikan rekomendasi dan informasi kepada para ASN untuk mengetahui kepangkatan yang tepat dengan latar belakang atau persyaratan yang telah dimiliki.

Suherman et al. (2021) dalam penelitiannya menggunakan data training sebanyak 103 siswa dan kemudian diperoleh data testing sebanyak 41 data siswa menghasilkan nilai akurasi sebesar 82,82%, penelitian ini dapat digunakan sebagai alternatif referensi bagi guru BK untuk dapat mengelompokkan siswa berdasarkan minat dan bakat. Tingkat akurasi data yang dihasilkan dengan menggunakan metode *data mining* C4.5 menghasilkan tingkat akurasi data yang cukup tinggi dan masuk dalam kategori sangat baik (Sari, 2021).

Penelitian yang dilakukan Monalisa & Hadi (2020) dalam penentuan jurusan baru dengan menggunakan data inputan 439 *record* yang berasal dari data rekapitulasi pendaftaran dan seleksi siswa kelas X. Setelah dilakukan penyeleksian dan pembersihan data diperoleh 360 *record*. Atribut yang digunakan dalam penelitian ini

adalah Jneis Kelamin, Nilai tes siswa seperti nilai Matematika, IPA, IPS dan Agama. Dengan menggunakan algoritma C4.5 diperoleh akurasi sebesar 87,22% dengan masing-masing nilai *precision/recall* sebesar 0,872/0,872 dan hasil klasifikasi didapatkan bahwa 71 data yang memiliki jurusan agama, 71 data agama diklasifikasi secara benar oleh algoritma C4.5. Selanjutnya dari 144 data yang memiliki jurusan IPS, 123 data diklasifikasikan secara benar, 20 data diklasifikasikan sebagai IPS, dan 1 data diklasifikasikan sebagai agama. Serta dari 146 data yang memiliki jurusan IPS, 120 data diklasifikasikan secara benar, 25 data diklasifikasikan sebagai IPA, dan 1 data diklasifikasikan sebagai Agama. Algoritma C4.5 sangat direkomendasikan untuk mengklasifikasikan penjurusan, dengan catatan jika atribut yang dimiliki dalam penentuan jurusan mayoritas kedalam bentuk numerik, karena Algoritma C4.5 memang lebih baik untuk menyelesaikan data-data yang bersifat numerik.

Berdasarkan penelitian sebelumnya, algoritma C4.5 telah berhasil diterapkan dalam dunia pendidikan yang diantaranya digunakan untuk penentuan jurusan dan penelusuran bakat dan minat siswa. Walaupun telah banyak penelitian yang dilakukan dalam mengolah data dengan menggunakan Algoritma C4.5, namun masih belum dapat diketahui dengan pasti teknik data mining dalam penentuan jabatan fungsional yang cocok untuk para aparatur sipil negara. Oleh karena itu, penulis mengangkat judul penelitian tentang **“Analisis dan Prediksi Perilaku Aparatur Sipil Negara dalam Menempati Jabatan Fungsional Menggunakan Algoritma C4.5”**.

## 1.2 Perumusan Masalah

Berdasar permasalahan yang ada, supaya tesis ini sesuai dengan tujuan yang ingin dicapai, maka penulis merumuskan beberapa permasalahan dalam tesis ini sebagai berikut:

1. Bagaimana membuat model aturan keterhubungan data ASN dalam menempati jabatan fungsional menggunakan algoritma C4.5?
2. Bagaimana menggunakan model aturan untuk mengklasifikasikan predikat akhir seorang aparatur sipil negara berdasarkan jenjang jurusan pendidikan?

### 1.3 Batasan Masalah

Agar pembahasan pada penelitian ini tidak menyimpang maka penulis membatasi ruang lingkup objek penelitian. Adapun ruang lingkup penelitian ini dibatasi dengan beberapa hal sebagai berikut:

1. Pengisian Jabatan Fungsional yang digunakan adalah perpindahan dari jabatan lain.
2. Algoritma teknik data mining yang digunakan adalah algoritma C4.5 untuk mendapatkan *decision tree*.
3. Data diperoleh dari database kepegawaian Pemerintah Kota Sungai Penuh.
4. Dalam membuat *rule* atau aturan, penulis menggunakan perangkat lunak Rapidminer studio 9.10 untuk melakukan analisis data.

### 1.4 Tujuan Penelitian

Tujuan yang ingin diperoleh dari penelitian ini agar lebih bermanfaat kedepannya adalah:

1. Mengetahui tingkat akurasi dari metode *Decision tree* dengan menggunakan Algoritma C4.5 dalam menganalisa dan memprediksi ASN untuk mengisi Jabatan Fungsional
2. Untuk mendapatkan factor-faktor yang berpengaruh terhadap perilaku ASN untuk mengisi Jabatan fungsional.

### 1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat kedepannya, yang beberapa diantaranya adalah:

1. Bagi pemerintah daerah, dapat memberikan referensi pembaharuan ilmiah sebagai informasi pendukung dalam mengambil kebijakan strategis dalam menganalisa Jabatan Fungsional di daerahnya.
2. Mempermudah pemerintah daerah dalam memaksimalkan informasi dalam memprediksi minat ASN untuk mengisi Jabatan Fungsional serta meningkatkan nilai akurasi yang jauh dari subjektivitas terhadap minat ASN untuk mengisi Jabatan Fungsional.

3. Bagi peneliti sendiri, bisa menambah khasanah dan variasi penelitian dalam penerapan teknik-teknik data mining pada area pemerintahan daerah.

## **1.6 Sistematika Penulisan**

Penelitian ini terbagi menjadi enam bab, dimana masing–masing bab terdiri dari beberapa sub–sub bab untuk menghasilkan pembahasan secara sistematis. Adapun enam bab tersebut ialah:

### **Bab I PENDAHULUAN**

Bab ini berisikan tentang latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

### **Bab II LANDASAN TEORI**

Bab ini berisikan tentang penerapan *Knowledge Discovery in Database (KDD)*, Pengertian *Data Mining*, klasifikasi, pengelompokan *Data Mining*, *Decision Tree*, Algoritma C4.5, ekstraksi aturan-aturan (*Rules*) dari *Decision Tree*, penelitian terdahulu yang relevan.

### **Bab III METODOLOGI PENELITIAN**

Bab ini berisi dasar penelitian, tahapan penelitian, sumber data, instrumen penelitian.

### **Bab IV ANALISA DAN PERANCANGAN**

Bab ini berisi tentang analisa permasalahan, analisa kebutuhan sistem, menerapkan sistem Algoritma C4.5, menentukan variabel dalam penghitungan hasil.

### **Bab V IMPLEMENTASI DAN HASIL**

Pada bab ini berisi pemaparan hasil penelitian yang digambarkan dalam bentuk uraian, tabel, grafik beserta pembahasannya.

### **Bab VI KESIMPULAN DAN SARAN**

Pada bab ini membahas tentang kesimpulan dari semua pembahasan yang ada dan saran-saran bagi para pembaca atau penelitian selanjutnya.