

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Data adalah sekumpulan fakta yang direpresentasikan ke dalam bentuk karakter baik huruf, angka, gambar dan lainnya yang dapat diproses menjadi sebuah informasi. Data juga dapat disebut sebagai “Bahan mentah” yang kemudian dapat diolah sehingga mendapatkan hasil produksi yang disebut “Bahan jadi” yaitu berupa informasi. Salah satu bentuk pengolahan data yaitu data mining. Data mining merupakan sebuah teknologi yang dapat memproses data menggunakan teknik dan metode tertentu dalam volume besar yang digunakan oleh perusahaan untuk mengubah data mentah menjadi informasi atau pengetahuan yang berguna untuk membuat suatu keputusan bisnis. Data mining disebut juga dengan *Knowledge Discovery in Database (KDD)* yang pada dasarnya mempunyai beberapa teknik yaitu Deskripsi, Klasifikasi, Klusterisasi, Asosiasi, dan Prediksi (T.Larose & D.Larose, 2005).

Klasifikasi merupakan teknik yang digunakan untuk menemukan model agar dapat menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek. Teknik klasifikasi ini dapat dilakukan untuk menentukan keputusan pada suatu perusahaan. Salah satunya yaitu untuk melakukan proses penilaian kinerja karyawan yang nantinya dapat digunakan sebagai acuan kenaikan jabatan. Kinerja karyawan yang baik sangat diperlukan untuk kesuksesan dan kemajuan sebuah perusahaan. Perusahaan yang memiliki karyawan berkualitas dan berinovasi dapat dengan mudah mengelola

perusahannya sehingga tujuan yang telah ditetapkan dapat tercapai. Kenaikan jabatan atau yang biasa disebut dengan promosi jabatan merupakan salah satu hal yang penting dalam meningkatkan kualitas, inovasi dan kinerja karyawan. Dengan adanya promosi jabatan, akan menambah semangat bekerja serta tanggung jawab karyawan. (Sunarti, 2019).

Proses penentuan karyawan yang akan mendapatkan promosi jabatan memiliki beberapa kriteria dan karakteristik penilaian berdasarkan standar yang telah ditentukan sebagai parameter untuk menentukan karyawan yang seperti apa yang akan memiliki potensi untuk promosi jabatan dan tidak berpotensi promosi jabatan. Untuk menentukan karyawan yang berpotensi promosi jabatan dapat menggunakan penerapan teknik data mining. Algoritma yang biasa digunakan untuk melakukan klasifikasi pada data mining diantaranya adalah *Naïve bayes*, *K-Nearest Neighbour*, *Support Vector Mechine*, *Artificial Neural Network*, dan *Decision Tree*.

*Decision Tree* merupakan metode klasifikasi yang cukup populer digunakan karena mudah untuk dipahami oleh manusia. Kelebihan dari metode ini adalah tingkat akurasi yang tinggi dan lebih mudah digunakan dibandingkan dengan algoritma lainnya (Suntoro, 2019). *Decision Tree* memiliki beberapa jenis algoritma diantaranya yaitu algoritma CART (Classification and Regression Tree), ID.3 C.45 dan C5.0 yang ditemukan oleh John Ross Quinlan pada tahun 1986. Algoritma C5.0 merupakan perluasan dari algoritma C4.5. Algoritma C5.0 lebih baik daripada C4.5 pada kecepatan, memori, dan efisiensi dan nilai akurasi yang lebih tinggi. Algoritma C5.0 cocok digunakan untuk kumpulan data besar (Kusrini & Taufiq Emha, 2009).

Seperti pada jurnal milik (Wahono & Riana, 2020) yang berjudul “Prediksi calon pendonor darah potensial dengan algoritma *Naïve Bayes*, *K-Nearest Neighbours* dan *Decision Tree C4.5*” dengan menggunakan data sebanyak 3710 orang, menunjukkan hasil akurasi sebesar 93.83% dengan *Decision Tree*, dan 85.15% dengan *Naïve Bayes*, dan 84.10% dengan algoritma *K-nearest Neighbours*. Pada penelitian lain yang berjudul “Prediksi pola kecelakaan kerja pada perusahaan non ekstraktif menggunakan algoritma *Decision Tree C4.5* dan *C5.0*” dengan menggunakan 200 data kecelakaan kerja menghasilkan jumlah pola yang berbeda, tergantung pada pembagian data training dan data testing yang digunakan. Semakin banyak data training yang digunakan, maka akan semakin tinggi hasil akurasi yang dihasilkan. Aturan atau *rules* sebagai representasi dari pola yang dihasilkan oleh algoritma C5.0 lebih ringkas daripada C4.5. Hasil keakuratan pola yang dihasilkan oleh Algoritma C4.5 lebih tinggi dibanding dengan Algoritma C5.0, yaitu 70%.

Berdasarkan latar belakang permasalahan diatas, penulis menyimpulkan bahwa implementasi algoritma *Decision Tree* memiliki tingkat ketepatan dan akurasi yang baik untuk melakukan klasifikasi dan prediksi dalam jumlah data yang cukup besar. Dengan penerapan dua algoritma yakni *Decision tree C4.5* dan *C5.0*, Maka penulis akan melakukan penelitian dengan judul “*Analisis Perbandingan Algoritma Decision Tree C4.5 dan C5.0 pada Data Karyawan Berpotensi Promosi Jabatan*” dengan tujuan untuk melihat perbandingan akurasi pada kedua algoritma tersebut serta memanfaatkan data dari Kaggle untuk pembuatan model dari algoritma tersebut. Pengambilan data pada penelitian ini didapatkan dari website Kaggle dengan jumlah data sebanyak 54.808 data.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah di uraikan, perumusan masalah yang akan dibahas yaitu:

1. Bagaimana penerapan algoritma *Decision Tree C4.5 dan C5.0* pada data karyawan berpotensi promosi jabatan?.
2. Bagaimana perbandingan tingkat akurasi algoritma *Decision Tree C4.5* dengan *C5.0* pada data karyawan yang berpotensi promosi jabatan?

## **1.3 Tujuan Penelitian**

Penelitian ini bertujuan untuk mengetahui perbandingan nilai akurasi dan rule atau aturan yang dihasilkan dari algoritma *Decision Tree C4.5 dan C5.0* dengan memanfaatkan dataset dari website kaggle terkait data karyawan yang berpotensi promosi.

## **1.4 Manfaat Penelitian**

Manfaat dari penelitian ini yaitu :

1. Dapat melakukan pengolahan data karyawan menjadi informasi yang baru dan unik untuk mengetahui potensi promosi jabatan.
2. Dapat mengetahui perbandingan nilai akurasi pada algoritma *Decision Tree C4.5 dan C5.0*.
3. Dapat mengetahui kriteria karyawan yang memiliki potensi promosi jabatan.
4. Sebagai referensi pada penelitian selanjutnya dengan penerapan algoritma *Decision Tree*.

## 1.5 Batasan Masalah

Ditentukannya beberapa batasan masalah agar dapat memfokuskan penelitian, maka ditetapkan batasan masalah sebagai berikut :

1. Data yang digunakan didapatkan dari website Kaggle dengan jumlah data sebanyak 54.808 data
2. Metode yang digunakan adalah *Decision Tree* dengan jenis *C4.5* dan *C5.0*
3. Tools yang digunakan adalah *RapidMiner* dan *Google Colab* dengan bahasa pemrograman *Python*.
4. Hasil atau visualisasi dari pengolahan data akan ditampilkan dalam bentuk model pohon keputusan, *rule* (aturan) dan tabel perbandingan tingkat akurasi.
5. Terdapat 14 atribut yang kemudian dilakukan proses seleksi atribut menggunakan metode backward elimination dengan hasil seleksi sebanyak 8 atribut yang akan digunakan.