

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining merupakan metode untuk memperoleh informasi berharga dari sekumpulan data, dimana proses dilakukan dengan menerapkan berbagai ilmu lain seperti matematika, statistika dan pengenalan pola. *Data Mining* dapat digunakan dan diterapkan dalam berbagai bidang, seperti bidang kedokteran, pemasaran, pendidikan dan bidang lainnya yang menggunakan data[1]. *Data mining* dapat digunakan untuk melakukan proses klasifikasi, prediksi, perkiraan dan memperoleh informasi yang bermanfaat dari sekumpulan data berukuran besar. Metode C4.5 merupakan suatu algoritma yang cocok untuk melakukan klasifikasi. Hasil dari klasifikasi dengan metode C4.5 yaitu sebuah pohon keputusan serta *rules* dari atribut yang dipakai[2].

Metode C4.5 adalah metode klasifikasi yang cukup terkenal untuk mengklasifikasikan data baik data yang bertipe diskrit maupun numerik. Pohon keputusan yang terbentuk dari klasifikasi dengan metode C4.5 dapat menjadi informasi baru dari atribut yang digunakan seperti bisa menemukan hubungan antar atribut ataupun *prediction*[3]. Algoritma C4.5 merupakan algoritma yang dimodifikasi dari algoritma ID3. *Input* dari algoritma ID3 berupa label *training*, sampel *training* dan atribut. Beberapa modifikasi yang dilakukan pada metode C4.5 seperti bisa mengatasi *missing value*, kontinuitas data, dan *pruning* [4]. Kelebihan utama dari algoritma C4.5 adalah mampu menghasilkan model berupa aturan yang mudah diinterpretasikan ataupun *tree*, tingkat akurasi yang cukup bagus, mampu memproses atribut dengan tipe numerik dan diskrit. Model yang dihasilkan dari proses “belajar” terhadap data pelatihan pada algoritma C4.5 adalah berupa sebuah pohon keputusan. Pohon keputusan inilah yang akan digunakan untuk melakukan klasifikasi, prediksi ataupun untuk hal lainnya[5].

Untuk mengetahui penerapan metode C4.5 dalam menyelesaikan berbagai kasus klasifikasi dalam kehidupan sehari-hari, maka disusun sebuah artikel penelitian dengan judul “**Penggunaan Metode C4.5 untuk Kasus Klasifikasi**”.

1.2 Perumusan Masalah

Adapun permasalahan dalam penelitian ini adalah:

1. Bagaimana menjabarkan cara kerja dari metode C4.5.
2. Bagaimana menjelaskan penerapan metode C4.5 untuk kasus klasifikasi.

1.3 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah untuk menjelaskan secara terperinci penggunaan metode C4.5 untuk kasus klasifikasi.

Adapun manfaat dari penelitian ini adalah:

1. Hasil rangkuman dapat dijadikan sebagai referensi dalam mempelajari dan memahami penggunaan metode C4.5 dalam kasus klasifikasi.
2. Mengetahui cara penerapan metode C4.5 dalam menyelesaikan kasus klasifikasi.

1.4 Pembatasan Masalah

Batasan dari penelitian ini meliputi:

1. Menjabarkan penggunaan metode C4.5 dalam menyelesaikan berbagai kasus klasifikasi.
2. Merincikan hasil penelitian yang diperoleh dari berbagai penerapan metode C4.5 tersebut.
3. Penelitian hanya membahas mengenai penerapan metode C4.5 yang telah pernah dilakukan sebelumnya.

1.5 Keterbaruan

1. Erwin Prasetyo dan Budi Prasetyo pada tahun 2020 melakukan penelitian untuk diagnosis penyakit jantung dengan metode C4.5 dan teknik bagging. Pada hasil penelitian didapat akurasi sebesar 72,98% dan dengan teknik bagging didapat akurasi sebesar 81,84%

2. Teguh Budi Santoso dan Dela Sekardiana pada tahun 2019 melakukan penelitian untuk menentukan kelayakan memberikan kredit pada koperasi dengan metode C4.5, Pada hasil penelitian didapat akurasi sebesar 97,5%
3. Alfian Faiz Izzulhaq dan Sulastri pada tahun 2020 melakukan penelitian untuk klasifikasi penjualan android. Pada hasil penelitian dengan 70% data training dapat akurasi sebesar 73,3%