

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining (Penambangan Data) merupakan sebuah teknik untuk menggali informasi tersembunyi untuk memperoleh manfaat lebih dari data yang ada. *Data mining* adalah proses memanfaatkan data dengan menggunakan teknik statistik, matematika, *artificial intelligence* dan *mechine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat. Adapun tujuan dari *data mining* sendiri adalah untuk melakukan klasifikasi, klasterisasi, menemukan pola asosiasi hingga melakukan peramalan (*predicting*)[1]. Istilah populer dari *data mining* lebih dikenal dengan sebutan KDD (*Knowledge Discovery from Database*) yang merupakan sekumpulan proses untuk menemukan pengetahuan yang bermanfaat dari suatu data[1].

Klasifikasi merupakan suatu proses dalam mencari suatu himpunan model (fungsi) untuk dapat mendeskripsikan dan membedakan kelas-kelas data atau konsep-konsep, yang memiliki tujuan agar dapat menggunakan model tersebut sehingga bisa memprediksi kelas dari objek yang belum diketahui kelasnya[2].

Mengenai dipilihnya algoritma C4.5 sebagai algoritma pengklasifikasian, karena tingkat akurasi algoritma ini dalam pengklasifikasian relatif tinggi, hal ini telah dibuktikan melalui jurnal peneliti yang dilakukan oleh (Khotimah and Istiawan 2018) yaitu tentang perbandingan algoritma C4.5, *Naïve Bayes* dan *K-Nearest Neighbour* untuk prediksi lahan kritis di kabupaten pemalang, dan hasil dari penelitian tersebut menyatakan bahwa algoritma klasifikasi C4.5 merupakan algoritma yang memiliki tingkat akurasi paling tinggi[3].

dataset Iris sendiri atau biasa disebut himpunan data bunga Iris merupakan kumpulan data bunga iris multivariat yang diperkenalkan oleh ahli statistik dan biologi Inggris Ronald Fisher dalam makalahnya tahun 1936. data terdiri dari 50 sampel dari masing-masing dari tiga spesies Iris (Iris setosa, Iris versi color dan Iris virginica) [4]. Empat fitur diukur dari masing-masing sampel yaitu panjang petal, lebar petal dan panjang sepal, lebar sepal. ukuran masing-masing data dalam satuan centimeter. Berdasarkan kombinasi keempat fitur ini, Fisher mengembangkan model diskriminan linier untuk membedakan spesies dari satu sama lain, dimana data ini atributnya berupa angka sehingga bisa dilakukan proses perhitungan klasifikasi.

Mengenai dipilihnya data iris sebagai objek karena Indonesia salah satu negara yang banyak akan sumber daya alamnya, maka dari itu Indonesia memiliki banyak ragam tumbuhan dan bunga yang tersebar di seluruh wilayah Indonesia, dari sekian banyak tumbuhan yang ada di Indonesia, hanya ada 20% yang sudah teridentifikasi [5]. Dan hanya beberapa atau sebagian masyarakat yang tau tentang bunga iris karena pembelajaran bunga iris di Indonesia masih menggunakan media cetak dan video [6] hal tersebut sulit dilakukan jika masyarakat ingin mengamati dan mengenali spesies bunga iris, karena setiap bunga iris memiliki kemiripan yang sama dari segi warna, bentuk dan ukuran sehingga sulit untuk mengetahui bunga iris termasuk ke kategori spesies yang mana, dengan adanya sistem ini maka diharapkan sistem dapat mengenali spesies bunga iris dengan memasukan nilai atribut dari panjang dan lebar sepal dan panjang dan lebar kelopak dalam ukuran sentimeter.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka permasalahan yang dapat dirumuskan adalah bagaimana analisis klasifikasi bunga iris menggunakan algoritma C4.5 ?

1.3 Batasan Masalah

Beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Penulis memfokuskan hanya menggunakan satu algoritma yaitu Algoritma C4.5
2. Untuk melihat akurasi algoritma tersebut pada data "Iris"
3. Data yang digunakan untuk klasifikasi yaitu dari dataset "Iris".

1.4 Maksud dan Tujuan Penelitian

Agar mendapat hasil klasifikasi algoritma C4.5 dengan sistem yang dibuat berupa tingkat akurasi dan waktu proses, serta dapat memprediksi jenis bunga pada klasifikasi untuk kasus data "Iris".

1.5 Manfaat Penelitian

1.5.1 Bagi peneliti

Dapat menambah pengetahuan tentang algoritma, khususnya algoritma C4.5 dan mendapat wawasan tentang spesies bunga iris. Selain itu peneliti mengetahui

tingkat akurasi dan cara penerapan algoritma C4.5 untuk klasifikasi pada dataset iris.

1.5.2 Bagi peneliti lain

Dapat mengetahui tingkat akurasi algoritma C4.5 khususnya dalam klasifikasi pada dataset iris serta penelitian ini dapat dikembangkan lebih lanjut dengan cara membandingkan beberapa algoritma klasifikasi lainnya.

1.6 Metode Penelitian

Tahap - tahap Metode penelitian yang akan dilakukan dalam penelitian ini adalah sebagai berikut :

1.6.1 Dasar Penelitian

Adanya penelitian ini karena adanya rumusan masalah yang ada, yang sudah terdapat pada rumusan masalah diatas yaitu untuk mendapat nilai akurasi dari algoritma C4.5 khususnya untuk study data iris dan menerapkan algoritma C4.5 ke sistem yang akan dibuat sehingga dapat mengklasifikasi spesies bunga iris.

1.6.2 Studi Literatur

Tahap ini adalah tahap pertama yaitu pencarian mengenai materi penelitian tentang klasifikasi *data mining*, algoritma C4.5, dan bunga iris. pencarian materi penelitian didasarkan pada penelitian - penelitian yang pernah ada yang berkaitan dengan penelitian yang dilakukan serta teori - teori yang sedang berkembang saat ini.

1.6.3 Pengumpulan Data

Pengumpulan data dilakukan dengan teknik data sekunder yaitu mengambil data yang sudah ada dipenelitian Ronald Fisher yang *publish* pada web *Machine Learning Repository*. "data berbentuk kuantitatif yaitu data yang terdiri dari kumpulan angka-angka hasil observasi" yang akan Diuji menggunakan metode klasifikasi dengan algoritma C4.5.

1.6.4 Pengolahan Data

Pada tahap pengolahan data akan dilakukan proses proses KDD (*Knowledge Discovery from Database*), sebelum data diimplementasikan ke sistem. lima langkah yang dilakukan untuk pemrosesan data adalah sebagai berikut (1) Pembersihan data (*data cleaning*), (2) Integrasi data (*data integration*), (*data cleaning*), (3) Seleksi data (*data selection*), (4) Transformasi data (*data transformation*) dan yang terakhir (5) penambangan data (*data mining*).

1.6.5 Perancangan dan Desain Sistem

Pada tahap ini, dilakukan perancangan terhadap sistem yang dibangun. Dengan melakukan perancangan umum sistem dan perancangan algoritma C4.5 ke sistem.

1.6.6 Implementasi

Langkah yang dilakukan pada tahap ini yaitu menerapkan algoritma C4.5 ke dalam sistem dengan menggunakan bahasa pemrograman PHP dengan memasukan file data teks Iris yang sudah di *preprocessing* ke sistem.

1.6.7 Pengujian

Pada tahap ini dilakukan pengujian dengan menggunakan algoritma C4.5 yang sudah diterapkan pada sistem dengan menggunakan study data iris sehingga akan mendapatkan hasil akurasi algoritma C4.5 pada sistem yang dibuat.

1.6.8 Hasil

Tahap ini adalah tahap terakhir yang dilakukan penulis yaitu mendapatkan hasil dari penelitian yang dilakukan, berupa tingkat akurasi algoritma dan sistem dapat memprediksi spesies bunga iris.

1.7 Sistematika Penulisan

Untuk penelitian ini terdiri dari lima bab:

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi tentang kajian pustaka yang mengenai penelitian terdahulu yang relevan dan dasar teori untuk dipakai guna menunjang penelitian.

BAB III ANALISIS DAN PERANCANGAN

Bab ini berisi tentang alat dan bahan penelitian serta alur penelitian, pada tahap ini membahas lebih lanjut tentang masing-masing dari metode penelitian satu – persatu.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini berisi tentang, pembuatan produk, implementasi algoritama yang digunakan, hasil dan akhir produk, serta hasil dan pembahasan.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan dan saran dari yang berhubungan dengan analisa dan optimalisasi berdasarkan penelitian yang telah dilakukan.

