

BAB I

PENDAHULUAN

1.1 Latar Belakang

Knowledge Discovery in Database (KDD) adalah suatu metode ekstraksi *non-trivial* yang tersirat suatu informasi tidak diketahui sebelumnya tetapi hasil yang diterima dari data tersebut terkandung potensi informasi (Subroto *et al.*, 2022). Proses yang dilakukan dalam KDD adalah *Data Mining*. merupakan cara mendapatkan pengetahuan (*Knowledge Discovery*) yang digali dari himpunan data dengan volume sangat besar, dan dapat mendukung dalam mengambil keputusan (C, 2024).

Algoritma C4.5 adalah algoritma yang digunakan untuk membuat *rule* dan suatu pohon keputusan. Algoritma C4.5 menggambarkan nilai atribut menjadi kelas yang dapat diterapkan untuk klasifikasi baru (Algoritma, *n.d.*).

Algoritma *Cart* merupakan algoritma yang umum dan banyak digunakan yang mengintegrasikan berbagai faktor dari sumber yang berbeda untuk masalah klasifikasi dan regresi berdasarkan *biner* rekursif (Amanda *et al.*, 2024). Algoritma ini tergolong dalam model *nonparametric* yang tidak memerlukan bentuk fungsional dan telah terbukti menjadi alat ampuh untuk masalah prediksi dan klasifikasi. *CART* mengandalkan partisi *biner* rekursif dari data dasar pembangunan pohon regresi. Pohon didirikan dari dataset yang dikumpulkan di simpul pohon akar dan setiap *node* dibagi menjadi dua *node* turun menggunakan variabel pemisahan. Pemilihan variabel pemisahan mencari penurunan variabilitas dalam *node* dan untuk peningkatan

variabilitas antara *node*, dengan setiap partisi yang diperoleh menghasilkan pohon dengan variabilitas yang lebih sedikit daripada pohon sebelumnya (Amanda *et al.*, 2024).

Kelapa sawit adalah tanaman tropis yang berasal dari Afrika Barat (*Economics et al.*, 2024). Kelebihan dari tanaman ini juga dapat ditanam diluar tempat asalnya, termasuk Indonesia. Tanaman ini telah banyak dibudidayakan dalam bentuk perkebunan dan pabrik di berbagai daerah di Indonesia (Gusmira *et al.*, 2024). Kelapa sawit merupakan tanaman industri yang digunakan sebagai bahan baku minyak nabati, minyak industri, dan bahan bakar. Kelapa sawit penting bagi Indonesia karena menciptakan lapangan kerja bagi masyarakat lokal dan merupakan sumber mata uang asing bagi Negara (Kurniawan *et al.*, 2023).

Tanaman kelapa sawit mulai berbunga dan membentuk buah setelah berumur 2-3 tahun (Tobing, 2015). Buah akan menjadi masak sekitar 5-6 bulan setelah penyerbukan. Proses pematangan buah kelapa sawit dapat dilihat dari perubahan warna kulit buahnya. Buah akan berubah menjadi merah jingga ketika masak. Pada saat buah masak, kandungan minyak pada daging buah telah maksimal. Jika terlalu matang, buah kelapa sawit akan jatuh dari tangkai tandannya (S *et al.*, 2024).

Ada beberapa tingkatan kematangan dari tandan buah segar (TBS) yang dipanen. Tingkat kematangan tersebut sangat mempengaruhi mutu panen, termasuk kualitas minyak sawit yang dihasilkan. Yaitu mentah dan matang saat matang tandan akan berwarna merah (Widjaja *et al.*, 2024). Untuk mengetahui tingkat kematangan TBS masih dilakukan dengan cara konvensional, yaitu dengan melihat perubahan warna dan jumlah brondolan yang jatuh. Saat buah telah dipanen, buah akan dikumpulkan sebelum diangkut truk ke pabrik (Rifqi & Suharjito, 2021). Buah yang layak untuk diangkut adalah buah yang telah matang, sementara buah mentah tidak

layak untuk diolah dan akan ditinggalkan di tempat. Tentunya hanya buah yang layak untuk diolah yang akan diterima dipembelian *Ramp* yang kemudian diangkut truk untuk dibawa ke pabrik (Rifqi & Suharjito, 2021).

Saat ini pembelian untuk TBS di *Ramp* dilakukan sortasi manual dimana sortasi dilakukan oleh petugas sortasi dan kemudian baru melakukan penentuan harga menurut prediksi. Maka dari itu, dibutuhkan keterlibatan teknologi yang dinamakan *Machine Learning* untuk melakukan klasifikasi terhadap kematangan TBS untuk penentuan harga, dalam hal ini digunakan sebuah metode pengolahan data yaitu *Data Mining*. *Data Mining* adalah cabang ilmu yang menggabungkan bidang-bidang ilmu komputer yang dimanfaatkan guna mencari pola dan informasi menarik dari sekumpulan data, atau disebut sebagai proses penguraian pengetahuan dalam *database* menggunakan metode tertentu, seperti kecerdasan buatan, *Machine Learning* dan statistik (Elfaladonna dkk,2021). Salah satu metode yang dapat diterapkan dalam *Data Mining* adalah klasifikasi (Swathy dkk, 2022). Klasifikasi adalah metode sederhana yang digunakan untuk mengenali kelas atau model data yang kemudian digunakan sebagai pendekatan dalam prediksi suatu permasalahan (Saputra dkk,2021). Metode klasifikasi yang diterapkan pada penelitian ini adalah algoritma C4.5 dan *CART*, keduanya adalah algoritma untuk membangun pohon keputusan atau *Decision Tree* (Syafrinal & Febrianti, 2023).

Berdasarkan penelitian Gangavarapu dkk tentang menganalisis kinerja prediksi *stroke* penelitian ini membandingkan algoritma *Decision Tree* C4.5 dan *CART* untuk mendapatkan hasil klasifikasi yang terbaik. Hasil terbaik didapatkan oleh algoritma C4.5 dengan akurasi sebesar 95,76%, sedangkan akurasi algoritma *CART* hanya 95,11% dengan pembagian data latih dan data uji sama besar, yakni 70:30. Hasil klasifikasi diuji dengan *RapidMiner* dalam mengklasifikasikan data. Tingkat

performance ditunjukkan dengan nilai akurasi. Nilai akurasi tersebut diperoleh dengan pengujian hasil klasifikasi terhadap data *training* dan data testing. Perbandingan nilai akurasi antar algoritma yang digunakan dapat diketahui algoritma terbaik dalam membuat klasifikasi data perikanan tangkap (Suryani *et al.*, 2022).

Penelitian juga dilakukan oleh Fiqih Aditiya dkk. Hasil yang diperoleh dalam memprediksi penjualan tempe dengan Algoritma C4.5 yaitu adanya 9 *rule* dengan tingkat akurasi mencapai 70% dengan nilai gain sebesar 0,433764175 (Aditiya *et al.*, 2022).

Penelitian terdahulu lainnya juga dilakukan oleh Vista Anestiviya dkk, didapatkan hasil perhitungan melalui perhitungan sistem dan manual didapatkan hasil untuk data a memiliki tingkat akurasi sebesar 100% dan untuk data b sebesar 80%, yang mengartikan bahwa C4.5 dapat disarankan untuk mengolah data siswa dalam hal membantu memberikan keputusan terbaik pemilihan jurusan siswa (Anestiviya *et al.*, 2021).

Berdasarkan penelitian berikutnya, dapat disimpulkan bahwa algoritma *Decision Tree* baik dengan maupun tanpa PSO, merupakan pilihan yang lebih baik daripada *CART* dalam konteks klasifikasi data pada studi ini. Meskipun penggunaan PSO tidak memberikan peningkatan yang signifikan dalam performa, algoritma *Decision Tree* tetap menunjukkan performa yang konsisten dan efektif dalam deteksi penyakit *stroke* (Fadlan *et al.*, 2024).

Penelitian ini melakukan klasifikasi terhadap dataset penelitian sebanyak 147 *record*. Dari sebanyak 147 *record* data dilakukan pengujian klasifikasi menggunakan algoritma *CART* sebanyak 5 kali yang dilakukan dengan pembagian data *training* menjadi 60%, 70%, 80%, 90%, dan 100%. Model pohon keputusan yang direkomendasikan pada penelitian ini adalah hasil dari pengujian data *training* 90%.

Pada pengujian data training 90% memberikan *root* akar awal pada pohon keputusan adalah tahun tanan pohon sawit yang selanjutnya mengarah ke fitur-fitur lainnya. Performa dari model pohon keputusan yang direkomendasikan pada penelitian ini berada pada nilai akurasi sebesar 96,21% (Sulistiawaty & Nurahman, 2023).

Berdasarkan uraian diatas maka akan dilakukan penelitian yang berjudul **“Analisis Perbandingan Algoritma C 4.5 dan Algoritma *CART* Pada Tandan Buah Sawit untuk Mengetahui Tingkat Kematangan (Studi Kasus di RAMP 789 Batang Peranap)”**.

1.2 Perumusan Masalah

Berdasarkan permasalahan yang dijabarkan pada latar belakang di atas, maka perlu dirumuskan permasalahan untuk memudahkan dalam proses pengolahan data nantinya serta merumuskan tujuan penelitian yang akan dirancang adalah sebagai berikut:

1. Bagaimana penerapan algoritma C 4.5 dan algoritma *CART* pada tandan buah sawit untuk mengetahui Tingkat kematangan pada RAMP 789 Batang Peranap?
2. Bagaimana Implementasi algoritma C 4.5 dan algoritma *CART* menggunakan *tools rapidminer* pada tandan buah sawit untuk mengetahui Tingkat kematangan pada RAMP 789 Batang Peranap?
3. Bagaimana menguji Tingkat akurasi dari algoritma C 4.5 dan algoritma *CART* pada tandan buah sawit untuk mengetahui Tingkat kematangan pada RAMP 789 Batang Peranap?

1.3 Batasan Masalah

Perlu adanya batasan masalah agar tidak terjadi penyimpangan dalam laporan penelitian ini, maka diterapkan batas-batas terhadap sistem yang akan diteliti, hal ini dimaksudkan agar langkah-langkah pemecahan masalah tidak menyimpang, adapun ruang lingkup penelitian diambil antara lain:

1. Data difokuskan pada data tingkat kematangan pada tandan buah sawit di *Ramp* 789 Batang Peranap.
2. Objek penelitian yaitu tandan buah sawit yang datanya didapat langsung di *Ramp* 789 Batang peranap.
3. Metode yang digunakan adalah algoritma C 4.5 dan algoritma *CART*.

1.4 Tujuan Penelitian

Didapatkan beberapa tujuan penelitian yang bisa dijadikan untuk menentukan hasil penelitian, dan tujuan penelitian adalah sebagai berikut:

1. Penerapan algoritma C 4.5 dan algoritma *CART* pada tandan buah sawit untuk mengetahui Tingkat kematangan pada RAMP 789 Batang Peranap.
2. Implementasi algoritma C 4.5 dan algoritma *CART* menggunakan *tools rapidminer* pada tandan buah sawit untuk mnegetahui Tingkat kematangan pada RAMP 789 Batang Peranap.
3. Menguji Tingkat akurasi dari algoritma C 4.5 dan algoritma *CART* pada tandan buah sawit untuk mengatahui Tingkat kematangan pada RAMP 789 Batang Peranap.

1.5 Manfaat Penelitian

Adapun manfaat penelitian yang akan dilakukan adalah sebagai berikut:

1. Membantu pemilik *Ramp* 789 untuk keakuratan dalam Tingkat kematangan.
2. Bagi peneliti selanjutnya dapat dijadikan sebagai bahan pertimbangan lebih lanjut, serta referensi terhadap penelitian yang sejenis.

1.6 Sistematika Penulisan

Dalam menyusun karya tulis ilmiah ini, agar dalam pembahasan terfokus pada pokok permasalahan dan tidak melebar kemasalah yang lain, maka penulis membuat sistematika penulisan karya tulis ilmiah sebagai berikut:

BAB I PENDAHULUAN

Dalam bab ini, penulis menguraikan tentang latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulis.

BAB II LANDASAN TEORI

Disini penulis akan menguraikan tentang kerangka teoritis yang berkaitan dengan pembuatan tesis dan bertujuan untuk menjelaskan mengenai prosedur, teori-teori, pengertian dan defenisi serta pendapat para ahli khususnya yang berhubungan dengan metode C 4.5 dan metode *CART*.

BAB III METODOLOGI PENELITIAN

Membahas mengenai kerangka kerja penelitian dan uraiannya sebagai berikut: penelitian pendahuluan, pengumpulan data, *study* literatur,

validasi data, analisa sistem, perancangan sistem, implementasi, pengujian hasil.

BAB IV ANALISA DAN PERANCANGAN

Menguraikan dan menerangkan mengenai tahapan-tahapan penyelesaian masalah, mendisain arsitektur sistem, cara kerja metode, perhitungan metode, serta hasil dari pohon keputusan.

BAB V IMPLEMENTASI DAN HASIL

Membahas mengenai implemtasi sistem, pengolahan data, dan tampilan hasil dari pohon keputusan.

BAB VI KESIMPULAN DAN SARAN

Memberikan kesimpulan dari keseluruhan pembahasan tesis mengenai hasil analisa, cara penggunaan aplikasi yang telah diimplementasikan dan saran kepada pemakai metode sertamenjadi referensi untuk masa mendatang.