

Komparasi Algoritma CART dan C 4.5 pada Citra Tandan Buah Sawit untuk Mengetahui Tingkat Kematangan dalam Penentuan Harga

Riris Agustin[✉], Sarjon Defit, Sumijan

Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK, Padang, 25221, Indonesia

ririsagustin014@gmail.com

Abstract

Oil palm is an industrial crop used as raw material for vegetable oil, industrial oil, and fuel. Oil palm is important for Indonesia because it creates jobs for local people and is a source of foreign currency for the country. Oil palm plants begin to flower and form fruit after 2-3 years. The fruit will ripen about 5-6 months after pollination. The ripening process of oil palm fruit can be seen from the change in the color of the fruit's skin. The fruit will turn orange-red when ripe. Based on this, this study aims to assess the maturity of an oil palm fruit bunch using the CART and C.4 5 methods. The CART and C.45 methods each have several stages that will produce entropy and gain values that will later form a decision tree. The dataset consists of 37 data consisting of 10 criteria from Ramp 789 Batang Peranap. Based on the implementation of the C4.5 algorithm and the CART algorithm, the level of maturity of oil palm fruit bunches on RAMP 789 Batang Peranap which produces an accuracy of 97.30%. These results were obtained based on Process data with testing using the RapidMiner application which produces a Decision tree that is useful as a reference for decisions in determining whether or not oil palm fruit bunches are ripe, which so far has only been done by prediction.

Keywords: Information Technology, Palm Fruit Bunches, Maturity Level, Cart, C 4.5

Abstrak

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. Tanaman kelapa sawit mulai berbunga dan membentuk buah setelah berumur 2-3 tahun. 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. Berdasarkan hal tersebut penelitian ini bertujuan untuk menilai kematangan sebuah tandan buah sawit dengan menggunakan metode CART dan C.4 5. Metode CART dan C.45 masing-masing memiliki beberapa tahapan yang akan menghasilkan nilai entropy dan gain yang nantinya akan membentuk suatu pohon keputusan. Dataset terdiri dari 37 data terdiri dari 10 kriteria yang berasal dari Ramp 789 Batang Peranap. Berdasarkan implementasi algoritma C4.5 dan algoritma CART, Tingkat kematangan tandan buah sawit pada RAMP 789 Batang Peranap yang menghasilkan akurasi sebesar 97,30%. Hasil tersebut didapat berdasarkan data Proses dengan pengujian menggunakan aplikasi RapidMiner yang menghasilkan pohon Keputusan yang berguna sebagai acuan untuk keputusan dalam menentukan masak atau tidaknya tandan buah sawit yang selama ini hanya dilakukan prediksi saja.

Kata kunci: Teknologi Informasi, Tandan Buah Sawit, Tingkat Kematangan, Cart, C 4.5

KomtekInfo is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



1. Pendahuluan

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 [1]. 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 [2]. Kelapa sawit adalah tanaman

tropis yang berasal dari Afrika Barat[3]. 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[4]. 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 [5]. Tanaman kelapa sawit mulai berbunga dan membentuk

buah setelah berumur 2-3 tahun. Buah akan menjadi masak sekitar 5-6 bulan setelah penyerbukan[6]. 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 [7]. 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, matang, dan terlewat matang[8]. 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. Buah yang layak untuk diangkut adalah buah yang telah matang dan terlewat matang, sementara buah mentah tidak layak untuk diolah dan akan ditinggalkan ditempat. Tentunya hanya buah yang layak untuk diolah yang akan diterima dipembelian Ramp yang kemudian diangkut truk untuk dibawa ke pabrik [9]. Saat ini pembelian untuk TBS diRAMP 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.

Metode klasifikasi yang diterapkan pada penelitian ini adalah algoritma C4.5 dan CART, keduanya adalah algoritma untuk membangun pohon keputusan atau decision tree. 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[10].

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. 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[11].

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 [12]. 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 [13].

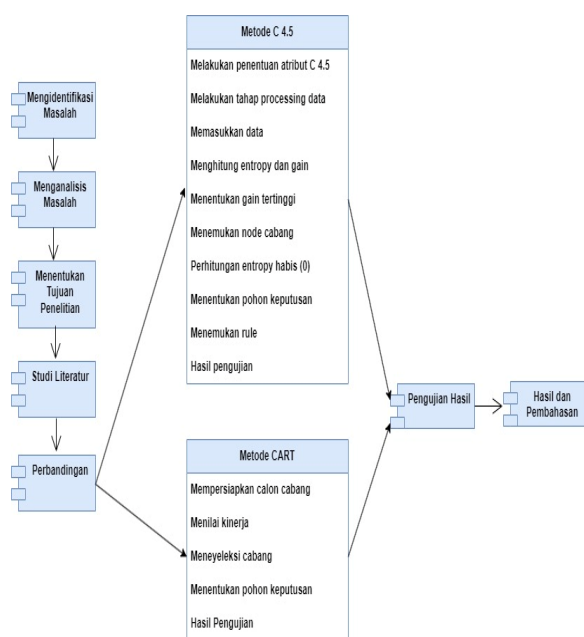
Penelitian terdahulu lainnya juga dilakukan oleh Vista Anestiviya dkk, didapatkan hasil perhitungan melalui perhitungan sistem dan manual didapatkan hasil untuk data a memiliki tingkat akurasi sebesar 100% dan untuk data b sebesar 80%, yang mengartikan bahwa C4.5 dapat disarankan untuk mengolah data siswa dalam hal membantu memberikan keputusan terbaik pemilihan jurusan siswa [14]. 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 [15]. 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 tanam 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% [16].

Berdasarkan literatur diatas meskipun penelitian sebelumnya sudah mengeksplorasi penggunaan metode cart dan c 4.5 dengan banyak bidang. Penelitian sebelumnya mungkin tidak secara spesifik dalam membahas tentang tandan buah sawit. Penelitian yang saat ini saya lakukan yaitu di tandan buah sawit dimana tandan buah sawit adalah hal yang perlu diteliti karna dapat memudahkan dalam penentuan kematangan. Penelitian ini bertujuan untuk menilai kematangan

sebuah tandan buah sawit dengan menghitung nilai entropy dan gain dari masing masing dataset yang sudah dikumpulkan atau didapatkan. Oleh karena itu, penelitian ini diharapkan tidak hanya melengkapi hasil penelitian sebelumnya, tetapi juga memberikan kontribusi baru yang lebih tepat sasaran.

2. Metodologi Penelitian

Dalam metodologi penelitian ini menjelaskan secara sistematis berguna sebagai pedoman untuk melakukan penelitian. Mulai dari mengidentifikasi masalah dan bahan penelitian yang sesuai dari jurnal untuk permasalahan yang diteliti. Bertujuan agar lebih terarah dan di harapkan bisa berjalan dengan baik sesuai dengan tujuan awal dari penelitian ini. Metodologi penelitian ini di gambarkan pada sebuah kerangka kerja dan peneliti akan menguraikan tahapan-tahapan dari kerangka kerja dapat dilihat pada Gambar 1.



Gambar 1 Kerangka kerja penelitian

Kerangka kerja penelitian adalah struktur atau rencana sistematis yang digunakan untuk memandu proses penelitian dari awal hingga akhir. Tujuan utamanya adalah untuk memastikan bahwa penelitian dilakukan dengan cara yang terorganisir, logis, dan dapat diterima secara ilmiah. Kerangka kerja ini membantu peneliti menetapkan langkah-langkah yang diperlukan dan memastikan bahwa semua aspek penting dari penelitian diperhitungkan. Kerangka kerja penelitian ini berfungsi sebagai panduan dan blueprint yang membantu peneliti tetap fokus dan terorganisir selama proses penelitian. Kerangka kerja Penelitian menguraikan tahapan-tahapan dari kerangka kerja penelitian sebagai berikut:

1. Mengidentifikasi Masalah

Mengidentifikasi masalah adalah langkah awal yang sangat penting dalam proses penelitian atau pemecahan masalah yang lebih umum. Proses ini melibatkan penentuan dan pemahaman dengan jelas tentang isu atau tantangan yang perlu dipecahkan atau diteliti. Proses mengidentifikasi masalah bertujuan untuk meninjau objek yang akan diteliti, menentukan ruang lingkup dan pencarian permasalahan pada objek penelitian. Mengidentifikasi masalah dengan tepat sangat penting untuk memastikan bahwa solusi yang diusulkan benar-benar efektif dan sesuai dengan kebutuhan yang ada. Ini membantu menghindari penyelesaian masalah yang tidak relevan atau tidak efektif.

2. Menganalisis Permasalahan

Menganalisis masalah adalah proses menyelidiki dan memahami masalah yang telah diidentifikasi untuk menentukan akar penyebabnya, dampaknya, dan solusi potensial. Tujuannya adalah untuk mendapatkan wawasan yang mendalam tentang masalah sehingga solusi yang diterapkan bisa lebih efektif dan tepat sasaran. Memahami permasalahan dalam ruang lingkup dan batasan masalah yang telah ditentukan dari tahapan mengidentifikasi masalah dimana menganalisis lebih ke melihat bagaimana permasalahan itu muncul, sehingga permasalahan akan lebih mudah dipahami. Menganalisis masalah secara menyeluruh membantu memastikan bahwa solusi yang diimplementasikan adalah yang paling efektif dan sesuai dengan kebutuhan. Ini juga membantu dalam mengurangi risiko dan biaya yang terkait dengan implementasi solusi yang tidak tepat.

3. Menentukan Tujuan dari penelitian

Setelah menganalisa permasalahan dan didapatkan rumusan permasalahan selanjutnya menentukan tujuan dari penelitian yang ingin di capai. Menentukan tujuan penelitian adalah proses menetapkan apa yang ingin dicapai melalui penelitian. Tujuan penelitian memberikan arah dan fokus untuk seluruh proses penelitian, mulai dari pengumpulan data hingga analisis dan penarikan kesimpulan. Tujuan ini harus jelas, spesifik, dan relevan dengan masalah yang diidentifikasi Tujuan penelitian harus berdasarkan dari apa rumusan masalah yang didapat baik itu untuk kepentingan pribadi dan kepentingan instansi. Menentukan tujuan penelitian dengan jelas dan terstruktur membantu memastikan bahwa penelitian tetap pada jalur yang benar dan fokus pada hasil yang diinginkan. Ini juga mempermudah perencanaan metodologi, pengumpulan data, dan analisis hasil.

4. Mempelajari Literatur

Mempelajari literatur, dalam konteks penelitian, adalah proses meninjau dan menganalisis karya-karya ilmiah dan sumber-sumber tertulis yang relevan dengan topik

atau pertanyaan penelitian Anda. Tujuannya adalah untuk memahami apa yang telah dipelajari sebelumnya, mengidentifikasi gap atau kekurangan dalam pengetahuan yang ada, dan membangun dasar teori atau kerangka kerja untuk penelitian. Mempelajari Literatur berguna untuk mempelajari dan memahami metode serta sebagai acuan untuk referensi untuk mendukung proses penelitian. Sumber literatur dalam penelitian ini dari jurnal-jurnal ilmiah terdahulu. Mempelajari literatur adalah bagian penting dari proses penelitian karena memberikan landasan teoritis, memperluas pemahaman tentang topik, dan membantu merumuskan pertanyaan penelitian serta metodologi yang tepat.

5. Mengumpulkan data-data yang dibutuhkan

Beberapa metode pengumpulan data yang dilakukan seperti Observasi yaitu Melakukan pengamatan data pada Ramp 789, yang bertujuan mengetahui permasalahan yang ada. Kemudian Wawancara, dalam melakukan wawancara peneliti mewawancarai owner ramp 789 mencari tau penyebab terjadinya permasalahan untuk di angkat menjadi topik penelitian ini. Selanjutnya yaitu Studi Pustaka, studi Pustaka ini merupakan pengumpulan bahan dan mempelajari jurnal-jurnal yang berhubungan dengan permasalahan penelitian ini.

6. Perbandingan Algoritma CART dan Algoritma C.45

Algoritma CART (Classification and Regression Trees) dan C4.5 adalah dua metode populer dalam pembelajaran mesin untuk membangun model pohon keputusan, masing-masing dengan karakteristik dan aplikasi yang berbeda. Pada proses Perbandingan ini dilakukan pembandingan antara algoritma cart dan algoritma C 4.5. Diawali dengan pengujian lalu nanti melihat hasil pengujian serta membandingkan kedua metode. Kedua algoritma memiliki kekuatan dan kelemahan masing-masing. CART lebih umum dan sederhana, cocok untuk tugas-tugas klasifikasi dan regresi, sedangkan C4.5 menawarkan pendekatan yang lebih canggih untuk klasifikasi dengan kemampuan penanganan data hilang dan atribut kontinu yang lebih baik. Pilihan antara keduanya sering kali tergantung pada jenis data dan kebutuhan spesifik dari tugas analisis yang dihadapi.

Tahapan Proses Algoritma C4.5. Menurut Lakshmi dan rekan-rekannya dalam karya mereka tentang algoritma C4.5, tahapan proses algoritma ini meliputi beberapa langkah utama untuk membangun pohon keputusan yang efektif. Algoritma C4.5, yang dikembangkan oleh Ross Quinlan, adalah perbaikan dari algoritma ID3 dan digunakan untuk klasifikasi data. Berikut adalah tahapan proses algoritma C4.5 berdasarkan literatur dan implementasi umumnya :

a. Menyiapkan data training

Menyiapkan data training untuk metode C4.5 adalah langkah penting dalam proses pembangunan model pohon keputusan. Tahap ini memastikan bahwa data yang digunakan untuk melatih model C4.5 dalam kondisi

optimal, sehingga model yang dihasilkan akurat dan efektif. Menyiapkan data dengan benar adalah langkah penting untuk memastikan bahwa model C4.5 dapat dilatih dengan efektif dan memberikan hasil yang akurat. Proses ini membantu meningkatkan kualitas model dan memastikan bahwa hasilnya dapat diandalkan untuk membuat prediksi atau keputusan berbasis data.

b. Menghitung nilai entropy.

Entropy merupakan perbedaan keputusan terhadap nilai atribut tertentu dengan parameter ketidakpastian. Semakin tinggi perbedaan keputusan maka semakin tinggi nilai entropy [14]. Dalam algoritma C4.5, menghitung nilai entropy adalah proses penting untuk menentukan seberapa baik suatu atribut dapat membagi data dalam hal mengurangi ketidakpastian atau ketidakpastian tentang kelas. Entropy adalah ukuran ketidakpastian atau ketidakhomogenan dari dataset. Nilai Entropy dihitung dengan Persamaan 1.

$$\text{Entropy}(A) = \sum_{z=1}^J -p_i * \log_2 p_i \quad (1)$$

Keterangan:

A : Himpunan Kasus

J : Jumlah parti A

Pi : Proporsi dari Ai terhadap A

Nilai entropy menunjukkan seberapa campur aduk kelas-kelas dalam dataset. Nilai yang lebih tinggi menunjukkan ketidakpastian yang lebih besar, sementara nilai yang lebih rendah menunjukkan bahwa data lebih homogen. Dalam C4.5, entropy digunakan bersama dengan gain informasi untuk memilih atribut terbaik untuk membagi dataset, dengan tujuan mengurangi entropy dan meningkatkan kemurnian subset data setelah pembagian.

c. Menghitung nilai gain.

Algoritma C4.5, gain informasi (information gain) adalah ukuran yang digunakan untuk menentukan seberapa baik sebuah atribut membagi data untuk mengurangi ketidakpastian. Gain informasi mengukur seberapa banyak informasi tentang kelas yang diperoleh dengan membagi data berdasarkan atribut tertentu. Gain merupakan tingkat pengaruh sebuah atribut terhadap keputusan atau ukuran efektifitas suatu variabel dalam mengelompokkan data. Gain informasi memberikan ukuran seberapa baik atribut memisahkan data berdasarkan kelas. Atribut dengan gain informasi yang lebih tinggi dianggap lebih baik karena memberikan pembagian yang lebih informatif dan mengurangi ketidakpastian dalam dataset. Gain dihitung dengan Persamaan 2.

$$\text{Gain}(A, W) = \text{Entropy}(A) - \sum_{z=1}^j \frac{|Az|}{|A|} * \text{Entropy } Si \quad (2)$$

Keterangan :

A : Himpunan Kasus

W : Atribut

J : Jumlah partisi atribut A

|Az| : Jumlah Kasus pada partisi ke-z

|A| : Jumlah kasus dalam A

Perhitungan gain informasi dalam C4.5 dimulai dengan menghitung entropi dataset awal untuk mengukur ketidakpastian kelas dalam dataset. Selanjutnya, data dibagi berdasarkan atribut yang dianalisis, dan entropi dihitung untuk setiap subset yang dihasilkan dari pembagian tersebut. Gain informasi dihitung sebagai selisih antara entropi awal dan rata-rata entropi tertimbang dari subset-subset tersebut. Atribut dengan gain informasi tertinggi dianggap yang terbaik untuk digunakan dalam pemisahan data karena memberikan pengurangan ketidakpastian terbesar.

Metode CART (Classification and Regression Trees) adalah teknik pembelajaran mesin yang digunakan untuk membangun pohon keputusan baik untuk klasifikasi maupun regresi. Tahapan metode CART meliputi pemilihan atribut dan kriteria pemisahan, pembangunan pohon keputusan, pemangkasan pohon untuk menghindari overfitting, dan evaluasi serta validasi hasil. Proses ini memastikan bahwa pohon keputusan yang dihasilkan efektif dalam klasifikasi atau regresi dengan performa yang baik pada data yang tidak terlihat sebelumnya.

1. Mempersiapkan calon cabang (Candidate split) merupakan langkah awal. Untuk menyusun daftar kandidat cabang mutakhir, persiapan menyeluruh dilakukan pada semua variabel prediktor.
2. Langkah kedua, menilai kinerja keseluruhan daftar kandidat cabang mutakhir dengan menghitung keseluruhan nilai kesesuaian.
3. Langkah terakhir, menyeleksi calon cabang yang akan digunakan, dengan menggunakan kriteria berdasarkan nilai kesesuaian tertinggi.

Pada tahap pengembangan pohon keputusan, nilai impurity digunakan saat pemilihan split atribut. Nilai impurity dapat ditentukan dengan menggunakan nilai Gini Index. Gini index diperlukan dalam penentuan titik pembelah terbaik (splitting optimal point). Semakin rendah nilai Gini index maka semakin tinggi ukuran kesamaannya. Gini index variabel T untuk dataset dengan m class diformulasikan pada Persamaan 3.

$$Gini(T) = 1 - \sum_{i=1}^m p_i^2 \quad (3)$$

Keterangan :

M : jumlah variabel kelas dalam dataset

Pi : frekuensi dari satu record dalam dataset yang mempunyai atribut kelas Ti

Gini Index mengukur ketidakmurnian dataset dan digunakan dalam CART untuk memilih atribut terbaik yang memisahkan data. Prosesnya melibatkan perhitungan Gini Index untuk dataset dan subset, serta perhitungan rata-rata tertimbang untuk menemukan atribut dengan pengurangan impurity terbesar. Perhitungannya dapat dilakukan dengan cara membagi jumlah variabel dalam class Ti dengan jumlah keseluruhan record dalam dataset. Penggunaan indeks gini terendah untuk membagi dataset menjadi dua bagian. Ketika data dipecah terhadap A menjadi dua menjadi dua himpunan bagian, D1 dan D2. gini index dapat diformulasikan pada Persamaan 4.

$$GiniA^{(D)} = \frac{|D1|}{|D|} Gini(D1) + \frac{|D2|}{|D|} Gini(D2) \quad (4)$$

Ketika sebuah atribut digunakan untuk membagi data, dataset dibagi menjadi beberapa subset berdasarkan nilai-nilai atribut tersebut. Gini Index dihitung untuk setiap subset dengan cara yang sama seperti untuk dataset keseluruhan. Misalnya, jika atribut A memiliki nilai D1t_1d1 dan d2d_Dt2, dataset dibagi menjadi dua subset St1S_{t_1}St1 dan St2S_{t_2}St2.

Gini Index untuk klasifikasi mengukur impurity dan digunakan untuk memilih atribut yang meminimalkan impurity dalam pohon keputusan. Mean Squared Error (MSE) untuk regresi mengukur variabilitas target dan digunakan untuk memilih atribut yang meminimalkan variabilitas dalam pohon keputusan. Gain atau pengurangan impurity dihitung sebagai selisih antara impurity awal dan impurity rata-rata tertimbang setelah pemisahan. Atribut dengan gain terbesar dipilih untuk pemisahan data lebih lanjut.

7. Pengujian

Setelah melakukan perhitungan manual maka akan dilakukan pengujian ke *software rapidminer*. Pengujian dilakukan untuk mengetahui tingkat kelayakan dari kedua metode. Pengujian dilakukan dengan menggunakan *software rapidminer* yang menghasilkan nilai akurasi masing masing.

8. Analisa Hasil

Setelah melakukan pengujian di *software RapidMiner*, analisis hasil pun diperoleh. Analisis ini didasarkan pada data yang telah diolah dan diinterpretasikan oleh *software*, memberikan wawasan mendalam mengenai performa model yang diuji. Berdasarkan hasil analisis tersebut, peneliti dapat menarik kesimpulan mengenai efektivitas dan keakuratan model, serta membuat keputusan yang tepat terkait temuan penelitian. Hasil analisis ini menjadi dasar untuk menyusun kesimpulan akhir dari penelitian, dengan mempertimbangkan hasil yang diperoleh dan relevansi temuan terhadap tujuan penelitian.

9. Hasil dan pembahasan

Setelah melakukan analisis hasil dari kedua metode, akan diperoleh kesimpulan yang menyajikan perbandingan efektivitas dan kinerja masing-masing metode. Langkah terakhir adalah menyusun hasil dan pembahasan, di mana hasil tersebut mencakup data yang telah diuji serta interpretasi mendalam mengenai temuan yang diperoleh. Pada tahap ini, pembahasan akan mengulas secara rinci tentang bagaimana hasil yang didapat sesuai dengan tujuan penelitian, membandingkan kelebihan dan kekurangan dari setiap metode, dan memberikan wawasan tentang implikasi praktis serta rekomendasi untuk langkah-langkah selanjutnya.

3. Hasil dan Pembahasan

Data alternatif ini berasal dari Di Ramp 789 Batang Peranap dan akan dievaluasi menggunakan metode C 4.5 dan algoritma CART untuk menilai kematangan buah. Evaluasi ini mencakup pengumpulan dan analisis data mengenai berbagai aspek layanan yang disediakan di lokasi tersebut. Metode C 4.5 dan algoritma CART digunakan untuk mengetahui tingkat kematangan yang dapat memberikan pemahaman yang mendalam mengenai faktor-faktor yang memengaruhi tingkat kematangan dari tandan buah sawit. Data alternatif ini didapat dari Di Ramp 789 Batang Peranap yang akan dievaluasi menggunakan metode C 4.5 dan algoritma CART untuk menilai kualitas layanan :

Tabel 1 Sampel Hasil Conversi Data Alternatif

Warna Kulit Tandan	Tingkat Kehilangan Brondolan (Loose Fruits)	Kekompakan Buah	Ukuran Tandan	Jumlah Buah Mentah	Kadar Air	Kondisi Fisik Tandan	Tingkat Keasaman (pH)	Tingkat Kematangan
Merah	Sedikit (0-10%)	Kompak	Sedang (10-20 kg)	Sedikit (0-10%)	Sedang (15-25%)	Utuh	Netral (4.5-5.5)	Matang
Hitam	Banyak (>30%)	Longgar	Kecil (<10 kg)	Banyak (>30%)	Tinggi (>25%)	Rusak	Asam (<4.5)	Mentah
Hitam	Sedang (10-30%)	Kompak	Besar (>20 kg)	Sedang (10-30%)	Rendah (<15%)	Utuh	Netral (4.5-5.5)	Mentah
Hitam	Banyak (>30%)	Longgar	Sedang (10-20 kg)	Sedang (10-30%)	Sedang (15-25%)	Rusak	Netral (4.5-5.5)	Mentah
Merah	Sedikit (0-10%)	Kompak	Besar (>20 kg)	Sedikit (0-10%)	Tinggi (>25%)	Utuh	Basa (>5.5)	Matang
Merah	Sedang (10-30%)	Kompak	Kecil (<10 kg)	Sedikit (0-10%)	Rendah (<15%)	Utuh	Netral (4.5-5.5)	Matang
Hitam	Sedikit (0-10%)	Longgar	Sedang (10-20 kg)	Banyak (>30%)	Sedang (15-25%)	Rusak	Asam (<4.5)	Mentah
Merah	Banyak (>30%)	Kompak	Sedang (10-20 kg)	Sedikit (0-10%)	Sedang (15-25%)	Utuh	Netral (4.5-5.5)	Matang
Hitam	Sedikit (0-10%)	Longgar	Kecil (<10 kg)	Sedang (10-30%)	Tinggi (>25%)	Rusak	Asam (<4.5)	Mentah
Merah	Banyak (>30%)	Longgar	Besar (>20 kg)	Banyak (>30%)	Rendah (<15%)	Rusak	Netral (4.5-5.5)	Matang
Merah	Sedikit (0-10%)	Kompak	Sedang (10-20 kg)	Sedikit (0-10%)	Sedang (15-25%)	Utuh	Netral (4.5-5.5)	Matang
Hitam	Sedang (10-30%)	Longgar	Kecil (<10 kg)	Banyak (>30%)	Tinggi (>25%)	Rusak	Asam (<4.5)	Mentah

1. Perhitungan C. 4 5

Pada proses C4.5, data alternatif harus dikonversi terlebih dahulu untuk memastikan kesesuaian dengan format yang dibutuhkan oleh algoritma. Data hasil konversi ini akan memungkinkan penerapan algoritma C4.5 dengan lebih efektif dalam analisis dan pengambilan keputusan. Dengan konversi data yang tepat, algoritma dapat menerapkan subkriteria dengan akurat, sehingga menghasilkan pohon keputusan yang lebih optimal dan sesuai dengan kebutuhan analisis yang dilakukan. Data hasil konversi dapat dilihat padat sebagai berikut :

Entropy Jumlah Buah Mentah

$$Entropy \text{ Sedikit (0-10\%)} = \left(-\frac{3}{17} * \log_2 \left(\frac{3}{17}\right)\right) + \left(-\frac{14}{17} * \log_2 \left(\frac{14}{17}\right)\right) = 0,6723$$

$$Entropy \text{ Sedang Sedang (10-30\%)} = \left(-\frac{8}{8} * \log_2 \left(\frac{8}{8}\right)\right) + \left(-\frac{0}{8} * \log_2 \left(\frac{0}{8}\right)\right) = 0$$

$$Entropy \text{ Banyak (>30\%)} = \left(-\frac{7}{12} * \log_2 \left(\frac{7}{12}\right)\right) + \left(-\frac{5}{12} * \log_2 \left(\frac{5}{12}\right)\right) = 0,9799$$

Entropy Kadar Air

$$Entropy \text{ Rendah (<15\%)} = \left(-\frac{3}{7} * \log_2 \left(\frac{3}{7}\right)\right) + \left(-\frac{4}{7} * \log_2 \left(\frac{4}{7}\right)\right) = 0,9852$$

$$\text{Entropy Sedang (15-25\%)} = \left(-\frac{5}{17} * \log_2\left(\frac{5}{17}\right) + \left(-\frac{12}{17} * \log_2\left(\frac{12}{17}\right)\right) = 0,8740$$

$$\text{Entropy Tinggi (>25\%)} = \left(-\frac{10}{13} * \log_2\left(\frac{10}{13}\right) + \left(-\frac{3}{13} * \log_2\left(\frac{3}{13}\right)\right) = 0,7793$$

Entropy Kondisi Fisik Tandan

$$\text{Entropy Utuh} = \left(-\frac{4}{19} * \log_2\left(\frac{4}{19}\right) + \left(-\frac{15}{19} * \log_2\left(\frac{15}{19}\right)\right) = 0,7425$$

$$\text{Entropy Rusak} = \left(-\frac{14}{18} * \log_2\left(\frac{14}{18}\right) + \left(-\frac{4}{18} * \log_2\left(\frac{4}{18}\right)\right) = 0,7642$$

Entropy Tingkat Keasaman (pH)

$$\text{Entropy Asam (<4.5)} = \left(-\frac{11}{11} * \log_2\left(\frac{11}{11}\right) + \left(-\frac{0}{11} * \log_2\left(\frac{0}{11}\right)\right) = 0$$

$$\text{Entropy Netral (4.5-5.5)} = \left(-\frac{5}{23} * \log_2\left(\frac{5}{23}\right) + \left(-\frac{18}{23} * \log_2\left(\frac{18}{23}\right)\right) = 0,7554$$

$$\text{Entropy Basa (>5.5)} = \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right) + \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) = 0,7793$$

Setelah didapatkan perhitungan untuk nilai Entropy, langkah berikutnya adalah proses pencarian Gain Informasi. Nilai Gain Informasi dihitung dengan mengukur pengurangan Entropy yang diperoleh setelah membagi dataset berdasarkan atribut tertentu. Proses ini melibatkan perhitungan selisih antara Entropy awal dataset dan rata-rata Entropy dari subset yang dihasilkan setelah pemisahan. Gain Informasi memberikan indikasi seberapa efektif atribut dalam mengurangi ketidakpastian dan meningkatkan kemurnian subset. Atribut dengan Gain Informasi tertinggi akan dipilih untuk pemisahan berikutnya dalam pohon keputusan, karena atribut tersebut memberikan kontribusi terbesar dalam meningkatkan kualitas klasifikasi. Proses ini dilakukan sebagai berikut :

Gain Warna Kulit Tandan :

$$0,9995 - ((18/37) * 0) - ((19/37) * 0,2975) = 0,8467$$

Gain Tingkat Kehilangan Brondolan :

$$0,9995 - ((15/37) * 0,9968) - ((8/37) * 0,9544) - ((14/37) * 0,9403) = 0,033$$

Gain Kekompakan Buah:

$$0,9995 - ((19/37) * 0,7425) - ((18/37) * 0,7642) = 0,2464$$

Gain Ukuran Tandan :

$$0,9995 - ((11/37) * 0,4395) - ((17/37) * 0,8740) - ((9/37) * 0,9183) = 0,2493$$

Gain Jumlah Buah Mentah :

$$0,9995 - ((17/37) * 0,6723) - ((8/37) * 0,000) - ((12/37) * 0,9799) = 0,3728$$

Gain Kadar Air :

$$0,9995 - ((7/37) * 0,9452) - ((17/37) * 0,8740) - ((13/37) * 0,7793) = 0,1377$$

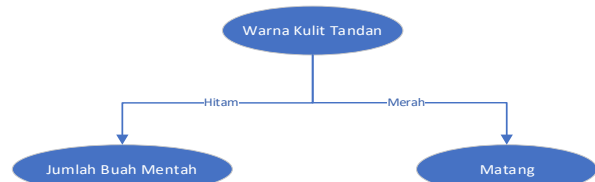
Gain Kondisi Fisik Tandan :

$$0,9995 - ((19/37) * 0,7425) - ((18/37) * 0,7462) = 0,2464$$

Gain Tingkat Keasaman (pH) :

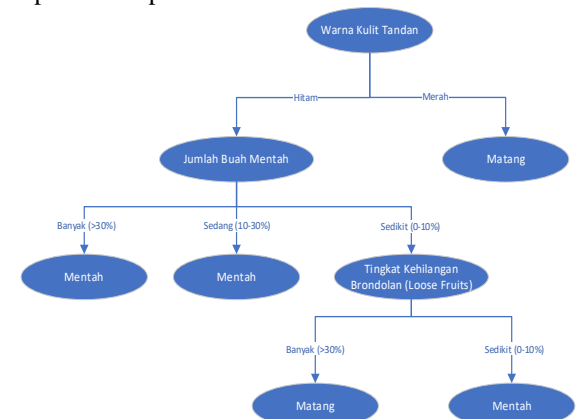
$$0,9995 - ((11/37) * 0,000) - ((23/37) * 0,7554) - ((3/37) * 0,9183) = 0,4555$$

Dari perhitungan diatas maka didapatkan pohon keputusan. Secara keseluruhan, pohon keputusan ini mengilustrasikan bagaimana data dibagi secara hierarkis berdasarkan atribut yang dipilih untuk mencapai keputusan akhir. Proses ini memastikan bahwa setiap cabang pohon memberikan informasi yang jelas dan spesifik, berdasarkan hasil perhitungan Gain Informasi dan Entropy ini menggambarkan struktur pengambilan keputusan pada note 1.



Gambar 2 Pohon Keputusan Note 1

Hasil gain tertinggi didapat dari Warna Kulit Tandan dengan warna hitam. Gambar tersebut menjelaskan dimana data yang akan kita poses adalah warna kulit tanda dengan warna hitam. Perhitungan selanjutnya dilakukan hingga menyajikan pohon keputusan yang dapat dilihat pada Gambar 3.



Gambar 3 Pohon keputusan C 4.5

Hasil pohon keputusan yang didapat dari variabel warna kulit tandan , jika warna kulit tandan merah maka tandan buah sawit dikatakan matang, jika warna kulit tandan hitam maka dilakukan pengujian pada jumlah buah mentah, jika jumlah buah mentah lebih dari 30% maka tandan buah sawit dikatakan mentah, jika jumlah buah mentah sedikit atau 10% maka tandan buah sawit dilakukan pengujian lagi ke variabel tingkat kehilangan brondolan, jika tingkat kehilangan brondolan lebih dari 30% maka tandan buah sawit dinyatakan matang dan jika tingkat kehilangan brondolan sedikit atau 10% maka tandan buah sawit dinyatakan mentah.

2. Perhitungan CART

Metode Classification and Regression Trees (CART) adalah algoritma untuk membangun pohon keputusan yang dapat digunakan untuk klasifikasi dan regresi. Berdasarkan algoritma CART akan ditentukan klasifikasi kumpulan data pada Tabel 1. Selanjutnya menentukan calon cabang kiri dan kanan yang membentuk cabang biner disajikan pada Tabel 2.

Tabel 2 Penentuan Calon Cabang Kanan dan Kiri

No	Calon Cabang Kiri	Calon Cabang Kanan
1	Warna Kulit Tandan Hitam	Warna Kulit Tandan Merah
2	Tingkat Kehilangan Brondolan (Loose Fruits) <= 10%	Tingkat Kehilangan Brondolan (Loose Fruits) > 10%
3	Tingkat Kehilangan Brondolan (Loose Fruits) <=30%	Tingkat Kehilangan Brondolan (Loose Fruits) > 30%
4	Kekompakan Buah Longgar	Kekompakan Buah Kompak
5	Ukuran Tandan <= 10 Kg	Ukuran Tandan > 10 Kg
6	Ukuran Tandan <= 20 Kg	Ukuran Tandan > 20 Kg
7	Jumlah Buah Mentah <= 10%	Jumlah Buah Mentah > 10%
8	Jumlah Buah Mentah <=30%	Jumlah Buah Mentah > 30%
9	Kadar Air <= 15%	Kadar Air > 15%
10	Kadar Air <=25%	Kadar Air > 25%
11	Kondisi Fisik Tandan Rusak	Kondisi Fisik Tandan Utuh
12	Tingkat Keasaman (pH) <= 4,5	Tingkat Keasaman (pH) <= 4,5
13	Tingkat Keasaman (pH) <= 5,5	Tingkat Keasaman (pH) <= 5,5

Setelah menentukan pembagian calon cabang kiri dan calon cabang kanan. Kemudian Mengitung banyak data dari setiap calon cabang kiri dan kanan yang dimana ini akan berfungsi untuk proses cart selanjutnya. Penghitungan ini penting untuk proses selanjutnya dalam metode CART, karena jumlah data di setiap cabang akan mempengaruhi evaluasi kesesuaian atribut dan pemilihan pemisahan terbaik. Pada tabel 2 hasil perhitungan dapat dilihat pada Tabel 3.

Tabel 3 Menghitung Data Calon Cabang Kanan dan Kiri

No	Banyak Data Calon Cabang Kiri (PL)	Banyak Data Calon Cabang Kanan (PR)
1	19	18

2	15	22
3	23	14
4	18	19
5	11	26
6	27	10
7	17	20
8	25	12
9	7	30
10	24	13
11	18	19
12	11	26
13	34	3

Setelah dilakukan perhitungan atau memisahkan data calon cabang kanan dan calon cabang kiri. Kemudian data dihitung banyaknya data calon cabang kanan dan calon cabang kiri. Proses perhitungan penilaian Keseluruhan calon cabang dengan menghitung nilai besaran kesesuaian dengan Persamaan 5.

$$PL = \frac{JUMLAH\ DATA\ CABANG\ KIRI}{JUMLAH\ DATA}$$

$$PR = \frac{JUMLAH\ DATA\ CABANG\ KANAN}{JUMLAH\ DATA} \tag{5}$$

Perhitungan dilakukan dengan cara menentukan kode PL (cabang kiri) dan PR (cabang kanan). Setelah penentuan kode, langkah berikutnya adalah menggunakan rumus untuk membagi jumlah calon cabang keseluruhan. Pembagian ini bertujuan untuk menghitung proporsi data di masing-masing cabang dan mengevaluasi efektivitas pemisahan berdasarkan atribut yang dipilih. Dengan memahami distribusi calon cabang ini, peneliti dapat melakukan analisis lebih lanjut untuk menentukan atribut yang paling optimal dalam membagi dataset dan meningkatkan akurasi model pohon keputusan. Untuk hasil perhitungan bisa dilihat pada tabel 4.

Tabel 4 Proses perhitungan penilaian Keseluruhan calon cabang

No	Banyak Data Calon Cabang Kiri	Banyak Data Calon Cabang Kanan
1	19 / 30 = 0,5135	18 / 30 = 0,4865
2	15 / 30 = 0,4054	22 / 30 = 0,5946
3	23 / 30 = 0,6216	14 / 30 = 0,3784
4	18 / 30 = 0,4865	19 / 30 = 0,5135
5	11 / 30 = 0,2973	26 / 30 = 0,7027
6	27 / 30 = 0,7297	10 / 30 = 0,2702
7	17 / 30 = 0,4594	20 / 30 = 0,5405
8	25 / 30 = 0,6757	12 / 30 = 0,3243
9	7 / 30 = 0,1892	30 / 30 = 0,8108
10	24 / 30 = 0,6486	13 / 30 = 0,3513
11	18 / 30 = 0,4865	19 / 30 = 0,5135
12	11 / 30 = 0,2973	26 / 30 = 0,7027
13	34 / 30 = 0,9189	3 / 30 = 0,08108

Setelah perhitungan nilai keseluruhan dengan pembagian banyaknya data. Maka didapat nilai gain dengan nilai gain tertinggi data cabang kiri yaitu 0,9189 dan nilai gain tertinggi data cabang kanan yaitu 0,8108. Proses berikutnya adalah Perhitungan P(j|tL) dan P(j|tR) yang dimana kita menghitung banyak data tingkat kematangan sawit. Adapun hasil tersebut dapat dilihat pada Tabel 5.

Tabel 5 P(j|tL) dan P(j|tR)

No	Keputusan	Jumlah Kategori j pada Noktah PL	Jumlah Kategori j pada Noktah PR
1	Matang	1	18
	Mentah	18	0
2	Matang	7	12
	Mentah	8	10
3	Matang	10	9
	Mentah	13	5
4	Matang	4	15
	Mentah	14	4
5	Matang	1	18
	Mentah	10	8
6	Matang	12	7
	Mentah	15	3
7	Matang	14	5
	Mentah	3	15
8	Matang	14	5
	Mentah	11	7
9	Matang	4	15
	Mentah	3	15
10	Matang	16	3
	Mentah	8	10
11	Matang	4	15
	Mentah	14	4
12	Matang	0	19
	Mentah	11	7
13	Matang	18	1
	Mentah	16	2

Tabel 5 merupakan hasil perhitungan $P(j|tL)$ dan $P(j|tR)$. Setelah didapat perhitungan banyak data tingkat kematangan sawit berdasarkan kriteria lalu kemudian proses perhitungan selanjutnya yaitu perhitungan $P(j|tL)$ dan $P(j|tR)$. Proses perhitungan Perhitungan $P(j|tL)$ dan $P(j|tR)$ adalah proses lanjutan yang dimana menggunakan persamaan yang dapat dilihat pada Persamaan 6.

$$P(j|tL) = \frac{\text{Jumlah Kategori } j \text{ pada Nota PL}}{\text{Jumlah Nota Pad}}$$

$$P(j|tR) = \frac{\text{Jumlah Kategori } j \text{ pada Nota PR}}{\text{Jumlah Nota Pad}} \quad (6)$$

Probabilitas kelas j diberikan bahwa sebuah data point berada di subset tLt_LtL (left) atau tRt_RtR (right) dari pemisahan yang dilakukan oleh sebuah atribut. Untuk setiap subset hasil pemisahan, hitung proporsi kelas j. Misalkan dataset asli memiliki dua subset setelah pemisahan: subset kiri tLt_LtL dan subset kanan tRt_RtR . $P(j|tL)$ adalah proporsi dari kelas j dalam subset kiri tLt_LtL . Sedangkan $P(j|tR)P(j|t_R)P(j|tR)$ adalah

proporsi dari kelas j dalam subset kanan tRt_RtR . Hasil tersebut dapat dilihat pada Tabel 6.

Tabel 6 Hasil Perhitungan $P(j|tL)$ DAN $P(j|tR)$

No	Keputusan	$P(j tL)$	$P(j tR)$
1	Matang	1 / 19 = 0,0526	18 / 18 = 1
	Mentah	18 / 19 = 0,9474	0 / 18 = 0
2	Matang	7 / 15 = 0,4667	12 / 22 = 0,5454
	Mentah	8 / 15 = 0,5333	10 / 22 = 0,4545
3	Matang	10 / 23 = 0,4347	9 / 14 = 0,6428
	Mentah	13 / 23 = 0,5652	5 / 14 = 0,3571
4	Matang	4 / 18 = 0,2222	15 / 19 = 0,7894
	Mentah	14 / 18 = 0,7778	4 / 19 = 0,2105
5	Matang	1 / 11 = 0,09091	18 / 26 = 0,6923
	Mentah	10 / 11 = 0,9091	8 / 26 = 0,3076
6	Matang	12 / 27 = 0,4444	7 / 10 = 0,7
	Mentah	15 / 27 = 0,5555	3 / 10 = 0,3
7	Matang	14 / 17 = 0,8235	5 / 20 = 0,25
	Mentah	3 / 17 = 0,1764	15 / 20 = 0,75
8	Matang	14 / 25 = 0,56	5 / 12 = 0,4166
	Mentah	11 / 25 = 0,44	7 / 12 = 0,5833
9	Matang	4 / 7 = 0,5714	15 / 30 = 0,5
	Mentah	3 / 7 = 0,4286	15 / 30 = 0,5
10	Matang	16 / 24 = 0,6666	3 / 13 = 0,2308
	Mentah	8 / 24 = 0,3333	10 / 13 = 0,7692
11	Matang	4 / 18 = 0,2222	15 / 19 = 0,7894
	Mentah	14 / 18 = 0,7777	4 / 19 = 0,2105
12	Matang	0 / 11 = 0	19 / 26 = 0,7308
	Mentah	11 / 11 = 1	7 / 26 = 0,2692
13	Matang	18 / 34 = 0,5294	1 / 3 = 0,3333
	Mentah	16 / 34 = 0,8888	2 / 3 = 0,6666

Setelah didapat perhitungan $P(j|tL)$ dan $P(j|tR)$ lalu kemudian dilakukan proses berikutnya yaitu Kesesuaian Untuk Calon Cabang Kesesuaian \emptyset (S|T). Proses berikutnya adalah Kesesuaian Untuk Calon Cabang Kesesuaian \emptyset (S|T) terbesar yang dimana ini adalah proses terakhir dari perhitungan CART yang dapat dilihat dengan Persamaan 7. Proses ini memastikan bahwa atribut yang dipilih untuk pemisahan memberikan keuntungan maksimal dalam hal pengurangan impurity atau kesalahan.

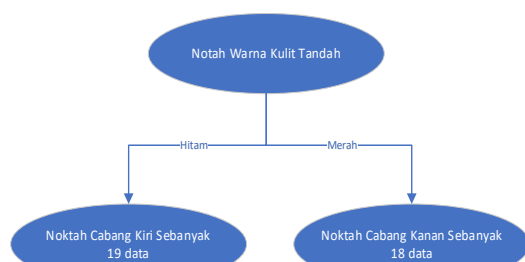
$$Q(S|T) = \frac{(\text{Nilai Max Keputusan} - \text{Nilai Min Keputusan})}{(\text{Nilai Max Keputusan} - \text{Nilai Min Keputusan})} \quad (7)$$

Notasi $\emptyset(S|T)$ dalam konteks ini bisa diartikan sebagai ukuran kesesuaian dari pemisahan dataset S berdasarkan kriteria tertentu, misalnya Gini Index atau Entropy. Kesesuaian $\emptyset(S|T)$ yang terbesar adalah atribut yang memberikan nilai gain tertinggi (Gini Gain, Information Gain, atau MSE Gain). Dalam prakteknya, pemilihan atribut dengan kesesuaian terbesar membantu dalam membangun pohon keputusan yang lebih efektif dengan pemisahan yang lebih baik, menghasilkan subset yang lebih murni atau memiliki kesalahan lebih rendah. Hasil tersebut dapat dilihat pada Tabel 7.

Tabel 7 Kesesuaian $\emptyset(S|T)$ terbesar

No	Q (S T)	2(PL*PR)	$\emptyset(S*T)$
1	1,8947368 42	0,49963477	0,946676406
2	0,1575757 58	0,482103725	0,07596786
3	0,4161490 68	0,470416362	0,195763331
4	1,1345029 24	0,49963477	0,566837107
5	1,2027972 03	0,417823229	0,502556611
6	0,5111111 11	0,394448503	0,201607012
7	1,1470588 24	0,496712929	0,569758948
8	0,2866666 67	0,438276114	0,125639153
9	0,1428571 43	0,30679328	0,043827611

Dalam pohon keputusan, kesesuaian $\emptyset(S|T)$ merujuk pada ukuran efektivitas pemisahan dataset berdasarkan atribut tertentu. Tujuan dari perhitungan ini adalah untuk menentukan atribut yang memberikan pemisahan terbaik untuk data, yang pada akhirnya akan membantu dalam membangun model keputusan yang lebih akurat dan efisien. Kesesuaian $\emptyset(S|T)$ terbesar menunjukkan atribut yang memberikan pemisahan terbaik untuk dataset, dengan mengurangi impurity atau kesalahan sebanyak mungkin. Dalam praktiknya, atribut dengan nilai kesesuaian tertinggi akan dipilih untuk membagi dataset pada setiap langkah pembuatan pohon keputusan, menghasilkan model yang lebih baik dan lebih akurat. Proses di atas dapat tercipta pohon keputusan yang dapat dilihat pada Gambar 5.



Gambar 5 Pohon Keputusan CART

Hasil pohon keputusan yang didapat menunjukkan dua aturan utama untuk klasifikasi. Pertama, rule untuk tingkat kematangan matang diperoleh dari 19 data yang terletak di node cabang kiri. Kedua, rule untuk tingkat kematangan mentah diperoleh dari 18 data yang terletak di node cabang kanan. Model ini secara efektif memisahkan data berdasarkan tingkat kematangan dengan menghasilkan dua subset yang jelas. Satu untuk data matang dan satu untuk data mentah. Dengan pemisahan yang tajam ini, pohon keputusan mampu mengklasifikasikan data dengan akurasi yang tinggi sesuai dengan kategori kematangan yang diinginkan.

4. Kesimpulan

Berdasarkan implementasi algoritma C4.5 dan algoritma CART dalam menentukan Tingkat kematangan tandan buah sawit pada RAMP 789 Batang Peranap yang menghasilkan akurasi sebesar 97,30%. Hasil tersebut didapat berdasarkan data Proses dengan pengujian menggunakan aplikasi RapidMiner. Pengujian ini menghasilkan pohon Keputusan yang berguna sebagai acuan untuk keputusan dalam menentukan masak atau tidaknya tandan buah sawit yang selama ini hanya dilakukan prediksi saja.

Daftar Rujukan

- [1] G. Subroto, N. Sulistiyowati, and A. A. Ridha, "Klasifikasi Jenis Kekerasan Pada Perempuan dan Anak Dengan Algoritma Multinomial Naïve Bayes," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 104–113, 2022, doi: 10.31539/intecomsv5i1.3598.
- [2] N. Suwaryo, A. Rahman, D. Marini, U. Atmaja, and A. Basri, "Klasterisasi Stok Produk Retail Untuk Menentukan Pergerakan Kebutuhan Konsumen Dengan Algoritma K-Means," *Bull. Inf. Technol.*, vol. 4, no. 2, pp. 306–312, 2023.
- [3] I. Economics, J. Volume, F. Ekonomi, and U. M. E-mail, "EVEKTIVITAS KERJASAMA INDONESIA DAN MALAYSIA," vol. 5, no. 1, pp. 25–37, 2024.
- [4] E. Gusmira, A. Ji, J. B. Km, and D. Muaro, "Literatur Review : Analisis Pengaruh Pepohonan Sawit Terhadap Suhu Disekitarnya Menggunakan Internet Of Things (IoT) Sensor Dh22 Eka Ariyani Safitri Internet of Things (IoT) adalah konsep di mana objek-objek fisik dilengkapi dengan Sensor DHT22 kelemb," vol. 3, no. 3, pp. 113–119, 2024.
- [5] R. Kurniawan, A. Halim, and H. Melisa, "Prediksi Hasil Panen Pertanian Salak di Daerah Tapanuli Selatan Menggunakan Algoritma SVM (Support Vector Machine)," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 2, pp. 903–912, 2023, doi: 10.30865/klik.v4i2.1246.
- [6] S. Y. L. Tobing, "Bab I ي ُحَض ُخ," *Galang Tanjung*, no. 2504, pp. 1–9, 2015.
- [7] D. S. N. S, U. Pembangunan, N. Veteran, J. Timur, M. R. N. Z, and S. Indira, "DAN BISNIS DI ERA DIGITAL Universitas Pembangunan Nasional ' Veteran ' Jawa Timur Universitas Pembangunan Nasional ' Veteran ' Jawa Timur," vol. 1, no. 5, pp. 8–16, 2024.
- [8] V. N. Widjaja, S. Tarmadja, and S. Gunawan, "Kehilangan hasil pada proses panen di perkebunan kelapa sawit lahan rendahan," *Kehilangan Has. pada proses panen di Perkeb. kelapa sawit lahan rendahan*, vol. 2, pp. 127–136, 2024.
- [9] M. Rifqi and Suharjito, "Deteksi Kematangan Tandan Buah Segar (Tbs) Kelapa Sawit Berdasarkan Komposisi Warna Menggunakan Deep Learning," *J. Tek. Inform. Atmaluhur*, vol. 6, no. 1, p. 40, 2021.

- [10] C. Algoritma, "KLASIFIKASI PENJUALAN WALMART MENGGUNAKAN," no. 02, pp. 122–129.
- [11] K. Amanda, D. Saripurna, and M. Z. Siambaton, "Penerapan Algoritma Cart dalam Penentuan Jurusan Siswa di SMA," *Hello World J. Ilmu Komput.*, vol. 2, no. 4, pp. 169–177, 2024, doi: 10.56211/helloworld.v2i4.404.
- [12] Suryani, D. Rahmadani, A. A. Muzafar, A. Hamid, R. Annisa, and Mustakim, "Analisis Perbandingan Algoritma C4.5 dan CART untuk Klasifikasi Penyakit Stroke," *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 197–206, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [13] P. Aditiya, U. Enri, and I. Maulana, "Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs Google Play Menggunakan Support Vector Machine," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1020, 2022, doi: 10.30865/jurikom.v9i4.4673.
- [14] V. Anestiviya, A. Ferico, and O. Pasaribu, "Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : Sman 1 Natar)," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 80–85, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTISI>
- [15] M. Fadlan *et al.*, "INFORMASI PADA MARKETPLACE Muhammad Fadlan," vol. 1, no. 3, pp. 560–565, 2024.
- [16] P. Sulistiawaty and Nurahman, "Klasifikasi Masa Awal Panen Sawit Pada PT. Mustika Sembuluh Menggunakan Algoritma Classification and Regression Tree," *J. TEKINKOM*, vol. 6, no. 2, p. 2023, 2023, doi: 10.37600/tekinkom.v6i2.946.