Journal of Artificial Intelligence and Software Engineering
Vol. 6, No. 2, Juni 2026, pp. 293~308
E-ISSN: 2777-001X, DOI: 10.30811/jaise.v6i2.9416

Classification of Taste Levels in Gerga Oranges Using the K-Cluster Classification Tree (K-CT) Method

Asep Syaputra[1*], Febriansyah[2], Buhori Muslim[3]
[1,2,3] Teknik Informatika, Institut Teknologi Pagar Alam, Kota Pagar Alam, 31521, Indonesia

Informasi Artikel

Diterima : 16 Juni 2026
Revisi : 20 Juni 2026
Publikasi : 30 Juni 2026

Kata Kunci:

Jeruk Gerga
Machine Learning
K-Cluster Classification Tree
K-Means
Classification Tree
Klasifikasi Rasa

ABSTRAK

Perkembangan teknologi telah mendorong pemanfaatan berbagai metode *machine learning* dalam mendukung pengambilan keputusan pada sektor pertanian dan hortikultura. Salah satu permasalahan yang masih sering ditemui adalah proses identifikasi tingkat rasa buah yang umumnya dilakukan secara subjektif berdasarkan pengamatan visual dan pengalaman individu, sehingga berpotensi menimbulkan ketidakkonsistenan dalam penilaian kualitas produk. Penelitian ini bertujuan untuk menerapkan metode *K-Cluster Classification Tree (K-CT)* dalam mengklasifikasikan tingkat rasa buah Jeruk Gerga berdasarkan karakteristik fisiknya. Metode *K-CT* merupakan pendekatan hibrida yang mengintegrasikan algoritma *K-Means Clustering* dan *Classification Tree* untuk meningkatkan kualitas model klasifikasi sekaligus mempertahankan efisiensi komputasi. Data penelitian berupa data primer yang diperoleh melalui observasi terhadap 700 sampel buah Jeruk Gerga yang berasal dari petani dan pedagang buah di Kecamatan Tanjung Sakti, Kabupaten Lahat. Setiap sampel direpresentasikan oleh enam atribut fisik, yaitu warna kulit, ukuran pori, keberadaan thrips, bentuk buah, tekstur kulit, dan diameter buah, dengan tingkat rasa sebagai variabel target. Data dibagi menjadi 80% data pelatihan dan 20% data pengujian. Hasil penelitian menunjukkan bahwa metode *K-CT* memperoleh akurasi klasifikasi sebesar 94,57%, lebih tinggi dibandingkan *Classification Tree*, *Random Forest,* dan *Gradient Boosting.* Selain itu, metode ini menunjukkan efisiensi waktu komputasi yang kompetitif serta mampu mengidentifikasi diameter buah sebagai atribut yang paling berpengaruh terhadap tingkat rasa Jeruk Gerga. Temuan ini menunjukkan bahwa *K-CT* memiliki potensi untuk diterapkan sebagai sistem pendukung keputusan yang objektif, akurat, dan efisien dalam klasifikasi kualitas buah.

ABSTRACT

Technological advancements have accelerated the adoption of various machine learning methods to support decision-making processes in the agricultural and horticultural sectors. One of the persistent challenges in fruit quality assessment is the identification of taste levels, which is commonly performed subjectively based on visual observation and individual experience. Such an approach may lead to inconsistencies in product quality evaluation. Therefore, this study aims to apply the K-Cluster Classification Tree (K-CT) method to classify the taste levels of Gerga oranges based on their physical characteristics. The K-CT method is a hybrid approach that integrates the K-Means Clustering algorithm with a Classification Tree to enhance classification performance while maintaining computational efficiency. The research utilized primary data collected through direct observation of 700 Gerga orange samples obtained from farmers and fruit traders in Tanjung Sakti District, Lahat Regency. Each sample was represented by six physical attributes, namely peel color, pore size, thrips presence, fruit shape, peel texture, and fruit diameter, while taste level served as the target variable. The dataset was divided into 80% training data and 20% testing data. The experimental results demonstrated that the K-CT method achieved a classification accuracy of 94.57%, outperforming Classification Tree, Random Forest, and Gradient Boosting models. Furthermore, the proposed method exhibited competitive computational efficiency and successfully identified fruit diameter as the most influential attribute affecting the taste level of Gerga oranges. These findings indicate that the K-CT method has

considerable potential to be implemented as an objective, accurate, and efficient decision support system for fruit quality classification.

Penulis Koresponden

Email: asepsyaputra68@itpa.ac.id

Cara sitasi IEEE::

[1] A. Sayaputra, F. Febriansyah, B. Muslim, "Classification of Taste Levels in Gerga Oranges Using the K-Cluster Classification Tree (K-CT) Method," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 6, no. 2, p. 293-308, Juni 2026. doi:10.30811/jaise.v6i2.9416

1. PENDAHULUAN

Perkembangan teknologi komputasi dan analisis data dalam beberapa dekade terakhir telah mendorong pemanfaatan teknologi kecerdasan buatan *(Artificial Intelligence)* dan *Machine Learning* pada berbagai bidang, termasuk sektor pertanian dan pangan [1]. Kemajuan teknologi tersebut memungkinkan proses pengolahan data dan pengambilan keputusan dilakukan secara lebih cepat, akurat, dan objektif berdasarkan pola yang diperoleh dari data. Dalam beberapa tahun terakhir, metode *Machine Learning* telah banyak diterapkan untuk mendukung berbagai aktivitas di sektor pangan, seperti identifikasi kualitas produk, deteksi cacat bahan pangan, prediksi karakteristik sensorik, serta klasifikasi tingkat kematangan dan kualitas buah [2]. Berbagai penelitian menunjukkan bahwa penerapan teknologi berbasis data mampu meningkatkan efisiensi proses pengelolaan dan evaluasi produk pangan, sekaligus mendukung upaya peningkatan kualitas hasil yang lebih konsisten dan terukur. Oleh karena itu, pengembangan model klasifikasi berbasis *Machine Learning* menjadi salah satu pendekatan yang penting untuk mendukung sistem penilaian kualitas pangan yang lebih efektif dan objektif [3].

Salah satu komoditas hortikultura yang memiliki nilai ekonomi tinggi di Indonesia adalah Jeruk Gerga. Jeruk Gerga dikenal sebagai salah satu varietas jeruk unggulan yang memiliki ukuran buah relatif besar, kandungan air yang tinggi, serta cita rasa yang khas. Komoditas ini banyak dibudidayakan di wilayah Sumatera dan menjadi salah satu sumber pendapatan masyarakat pada sektor pertanian. Tingginya permintaan pasar terhadap buah dengan kualitas rasa yang baik menyebabkan proses identifikasi karakteristik rasa menjadi aspek yang penting dalam rantai distribusi dan pemasaran produk. Akan tetapi, proses penentuan rasa buah umumnya masih dilakukan secara manual melalui pengamatan visual atau pengujian langsung oleh konsumen maupun pedagang [4]. Pendekatan tersebut memiliki kelemahan karena bersifat subjektif dan sering menghasilkan penilaian yang tidak konsisten.

Penentuan tingkat rasa buah berdasarkan karakteristik fisik merupakan permasalahan yang menarik untuk dikaji dalam bidang *Machine Learning*. Secara umum, karakteristik fisik seperti warna kulit, ukuran pori, bentuk buah, tekstur kulit, keberadaan bercak akibat serangan hama, dan ukuran buah sering digunakan sebagai indikator untuk memperkirakan kualitas buah [5]. Namun demikian, hubungan antara karakteristik fisik dan tingkat rasa tidak selalu bersifat linier sehingga diperlukan metode analisis yang mampu mempelajari pola hubungan tersebut secara efektif. Oleh karena itu, pengembangan model klasifikasi berbasis data menjadi salah satu solusi yang dapat digunakan untuk membantu proses identifikasi rasa buah secara lebih objektif [6].

Penelitian terdahulu menunjukkan bahwa metode berbasis pohon keputusan memiliki kemampuan yang baik dalam menyelesaikan berbagai permasalahan klasifikasi pada bidang pertanian dan pangan. Penelitian yang dilakukan oleh Ok, Akar, dan Gungor menggunakan metode Random Forest untuk klasifikasi lahan pertanian berdasarkan citra multispektral *SPOT-5*. Hasil penelitian menunjukkan bahwa metode Random Forest mampu menghasilkan akurasi keseluruhan *(overall accuracy)* sebesar 85,89%, yang lebih tinggi dibandingkan metode *Maximum Likelihood Classification (MLC).* Temuan tersebut menunjukkan bahwa pendekatan ensemble berbasis pohon keputusan memiliki kemampuan yang baik dalam menangani data pertanian yang kompleks dan multivariat [7].

Penelitian lain yang dilakukan oleh Utamima et al. mengkaji klasifikasi varietas beras menggunakan beberapa varian metode *Gradient Boosting*. Hasil eksperimen menunjukkan bahwa algoritma *Light Gradient Boosting Machine (LightGBM)* memperoleh tingkat akurasi sebesar 98,14%, yang merupakan performa terbaik dibandingkan metode *boosting* lainnya. Hasil tersebut menunjukkan bahwa pendekatan *boosting* mampu

meningkatkan kemampuan prediksi melalui proses pembelajaran iteratif yang memperbaiki kesalahan klasifikasi pada model sebelumnya [8].

Selain itu, penelitian yang dilakukan oleh Awate dan Nagne pada klasifikasi tanaman menggunakan citra Sentinel-2 menunjukkan bahwa *Random Forest* mampu menghasilkan akurasi validasi sebesar 97,91% dan mengungguli *Support Vector Machine* yang memperoleh akurasi 95,83%. Hasil penelitian tersebut memperkuat bahwa metode berbasis pohon keputusan mampu memberikan performa klasifikasi yang tinggi pada data pertanian yang memiliki karakteristik *heterogen* [9].

Berdasarkan berbagai penelitian tersebut dapat disimpulkan bahwa metode berbasis pohon keputusan, baik dalam bentuk *Decision Tree*, *Random Forest,* maupun *Gradient Boosting,* telah banyak digunakan untuk menyelesaikan berbagai permasalahan klasifikasi pada bidang pertanian dan pangan dengan tingkat akurasi yang tinggi. Meskipun demikian, penelitian yang secara khusus mengkaji klasifikasi tingkat rasa buah Jeruk Gerga *(Citrus nobilis Lour. var. microcarpa)* menggunakan pendekatan *K-Cluster Classification Tree (K-CT)* masih sangat terbatas. Selain itu, sebagian besar penelitian terdahulu berfokus pada proses klasifikasi secara langsung tanpa mempertimbangkan pembentukan kelompok data yang lebih homogen sebelum pembentukan model klasifikasi. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut dengan mengintegrasikan proses klasterisasi dan klasifikasi melalui metode *K-Cluster Classification Tree (K-CT)* guna meningkatkan kemampuan identifikasi tingkat rasa buah berdasarkan karakteristik fisik yang diamati.

Pada sisi lain, algoritma *K-Means* merupakan salah satu metode *clustering* yang banyak digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik. Penggunaan *clustering* sebelum proses klasifikasi memungkinkan pembentukan kelompok data yang lebih homogen sehingga model klasifikasi dapat dibangun secara lebih efektif [10]. Beberapa penelitian menunjukkan bahwa kombinasi antara proses *clustering* dan klasifikasi mampu meningkatkan kualitas model prediktif pada berbagai kasus. Namun demikian, penelitian yang mengintegrasikan algoritma *K-Means Clustering* dengan *Classification Tree* untuk mengidentifikasi tingkat rasa buah masih relatif terbatas, khususnya pada komoditas Jeruk Gerga.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan penerapan metode *K-Cluster Classification Tree (K-CT*) untuk mengklasifikasikan tingkat rasa buah Jeruk Gerga berdasarkan karakteristik fisiknya. Metode *K-CT* merupakan pendekatan hibrida yang mengombinasikan proses pengelompokan data menggunakan *K-Means* dengan pembentukan model klasifikasi menggunakan *Classification Tree*. Melalui pendekatan ini, data terlebih dahulu dikelompokkan ke dalam beberapa klaster berdasarkan tingkat kemiripan atributnya, kemudian pada setiap klaster dibangun model *Classification Tree* secara independen. Pendekatan tersebut diharapkan mampu menghasilkan model yang lebih akurat sekaligus mempertahankan efisiensi komputasi dibandingkan metode klasifikasi pohon konvensional.

Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh melalui observasi terhadap 700 sampel buah Jeruk Gerga yang berasal dari Kecamatan Tanjung Sakti, Kabupaten Lahat. Setiap sampel direpresentasikan menggunakan atribut warna kulit, ukuran pori, keberadaan thrips, bentuk buah, tekstur kulit, diameter buah, dan tingkat rasa sebagai variabel target. Selanjutnya, performa metode *K-CT* akan dibandingkan dengan beberapa metode berbasis pohon lainnya, yaitu *Classification Tree (CT), Random Forest* (RF), dan *Gradient Boosting (GB),* menggunakan parameter evaluasi berupa *Accuracy*, *Precision*, *Recall*, *F1-Score*, *True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR),* dan *False Negative Rate* (FNR).

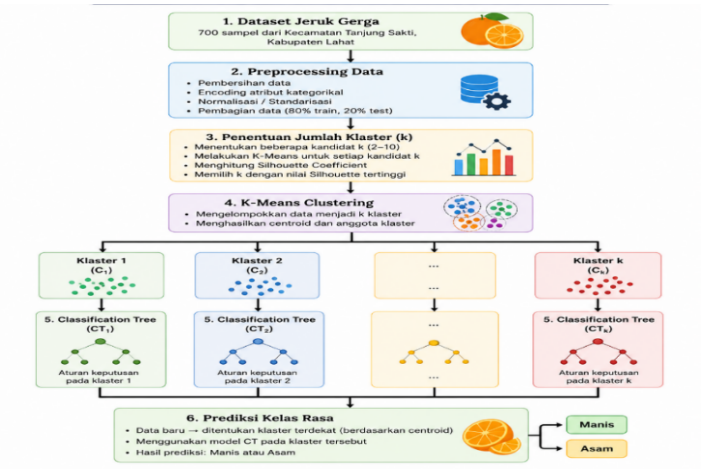
2. METODE

2.1. Metode *K-Cluster Classification Tree (K-CT)*

Metode *K-Cluster Classification Tree (K-CT)* merupakan pendekatan klasifikasi hibrida yang mengintegrasikan algoritma *K-Means Clustering* dengan *Classification Tree*. Integrasi kedua metode tersebut bertujuan untuk meningkatkan kemampuan model dalam mengidentifikasi pola data yang kompleks sekaligus mengurangi beban komputasi pada proses pembentukan pohon keputusan. Pada metode *Classification Tree* konvensional, seluruh data pelatihan digunakan secara langsung untuk membangun struktur pohon keputusan [11]. Pendekatan tersebut dapat menyebabkan peningkatan kompleksitas komputasi ketika jumlah data semakin besar. Oleh karena itu, penelitian ini menerapkan proses klasterisasi sebagai tahap awal untuk membagi data ke dalam kelompok-kelompok yang lebih homogen sebelum dilakukan proses klasifikasi. Secara umum, mekanisme metode *K-CT* terdiri atas empat tahapan utama, yaitu:

1. *Data preprocessing*.
2. Pembentukan klaster menggunakan algoritma *K-Means*.
3. Pembangunan model *Classification Tree* pada setiap klaster.

4. Prediksi kelas menggunakan model yang sesuai dengan klaster data baru.0-



Gambar 1. Arsitektur Metode *K-CT*

Gambar 1 memperlihatkan alur kerja metode *K-Cluster Classification Tree (K-CT)* yang diusulkan dalam penelitian ini. Proses dimulai dengan tahap data preprocessing untuk memastikan seluruh atribut yang digunakan berada dalam format yang sesuai dan siap diproses oleh algoritma *Machine Learning*. Tahap ini mencakup pemeriksaan kualitas data, pengkodean atribut kategorikal, serta penyesuaian struktur data guna meminimalkan potensi inkonsistensi yang dapat memengaruhi kinerja model [12]. Kualitas data yang baik menjadi faktor penting karena berkontribusi terhadap stabilitas dan akurasi hasil klasifikasi. Setelah *preprocessing*, dilakukan penentuan jumlah klaster optimal menggunakan metode *Silhouette Score.* Metode ini digunakan untuk mengevaluasi kualitas hasil pengelompokan berdasarkan tingkat kesamaan data dalam satu klaster dan perbedaannya dengan klaster lain [13]. Hasil klasterisasi kemudian digunakan sebagai dasar pembentukan model *Classification Tree* pada masing-masing klaster sehingga proses pembelajaran dilakukan pada kelompok data yang lebih homogen. Pendekatan ini memungkinkan pola hubungan antara atribut fisik buah dan tingkat rasa dipelajari secara lebih efektif. Tahap akhir menghasilkan klasifikasi tingkat rasa buah Jeruk Gerga yang selanjutnya digunakan untuk mengevaluasi performa model dan mengidentifikasi atribut yang paling berpengaruh terhadap tingkat rasa buah.

2.2. Representasi Data

Misalkan tersedia himpunan data pelatihan:

$$X = \{x_1, x_2, \dots, x_m\} \quad (1)$$

X : Dataset penelitian
m : Jumlah sampel
x_i : Sampel ke-i
Setiap sampel memiliki label kelas:

$$Y = \{y_1, y_2, \dots, y_m\} \quad (2)$$

Pada penelitian ini yang menunjukkan kategori tingkat rasa buah Jeruk Gerga.:

$$y_i \in \{Manis, Asam\} \quad (3)$$

yang menunjukkan kategori tingkat rasa buah Jeruk Gerga.

2.3. Tahap *Clustering* Menggunakan *K-Means*

Tahapan pertama dalam metode *K-CT* adalah melakukan pengelompokan data menggunakan algoritma *K-Means*. Tujuan utama proses ini adalah menghasilkan kelompok data yang memiliki karakteristik serupa sehingga model klasifikasi yang dibangun pada setiap klaster dapat bekerja lebih efektif.
Fungsi objektif *K-Means* dinyatakan sebagai:

$$J = \sum_{i=1}^{k} \sum_{x_j \in C_i} || x_j - c_i ||^2 \quad (4)$$

J : Fungsi objektif
k : Jumlah klister
C_i : Klaster ke-i
c_i : Centroid klaster ke-i

Nilai fungsi objektif yang semakin kecil menunjukkan kualitas klaster yang semakin baik karena jarak antaranggota klaster terhadap centroid semakin dekat.

2.4. Penentuan Jumlah Klaster Optimal

Pemilihan jumlah klaster merupakan faktor penting dalam proses klasterisasi. Pada penelitian ini, jumlah klaster optimal ditentukan menggunakan *Silhouette Coefficient*.
Nilai *silhouette* dihitung menggunakan persamaan:

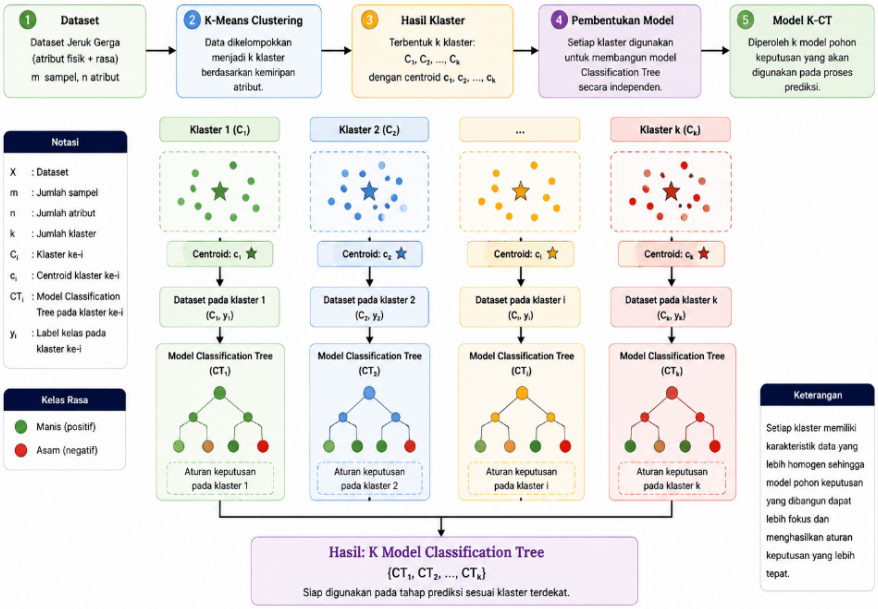
$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (5)$$

$a(i)$: Jarak rata-rata intra-klaster
$b(i)$: Jarak rata-rata ke klaster terdekat
$s(i)$: Nilai silhouette

Nilai *silhouette* berada pada rentang −1 hingga 1. Semakin mendekati nilai 1 menunjukkan bahwa suatu data berada pada klaster yang tepat.

2.5 Pembentukan Model *Classification Tree*

Setelah proses klasterisasi selesai dilakukan, setiap klaster digunakan sebagai dataset independen untuk membangun model *Classification Tree*. Pendekatan ini memungkinkan setiap model mempelajari karakteristik lokal yang lebih spesifik dibandingkan model yang dibangun menggunakan keseluruhan data.



Gambar 2. Struktur Pembentukan Model *K-CT*

Gambar 2 menunjukkan bahwa setiap klaster menghasilkan satu model *Classification Tree* yang berbeda. Dengan demikian, proses pembelajaran dilakukan secara terpisah sesuai karakteristik data pada masing-masing klaster. Strategi tersebut memungkinkan model memperoleh pola klasifikasi yang lebih spesifik dan meningkatkan kemampuan generalisasi pada data baru.

$$C_1, C_2, \ldots, C_k \quad (6)$$

Rumus diatas merupakan himpunan klaster yang dihasilkan. Model klasifikasi untuk setiap klaster dinyatakan sebagai berikut:

$$\hat{y}_i = f_{CT_i}(C_i) \quad (7)$$

2.6 Pemilihan Atribut pada Classification Tree

Pada penelitian ini, proses pemilihan atribut pada algoritma *Classification Tree* dilakukan menggunakan kriteria *Gini Impurity*. Kriteria ini digunakan untuk mengukur tingkat ketidakmurnian *(impurity)* suatu node berdasarkan distribusi kelas pada data yang berada di dalam node tersebut [14]. Semakin kecil nilai Gini Impurity, semakin homogen data pada node dan semakin baik kemampuan atribut dalam membedakan kelas target. Nilai *Gini Impurity* dihitung menggunakan Persamaan berikut:

$$Gini = 1 - \sum_{i=1}^{c} p_i^2 \quad (8)$$

p_i : Probabilitas kelas ke-i
c : Jumlah kelas

Atribut yang menghasilkan nilai impuritas terkecil dipilih sebagai pemisah node karena memiliki kemampuan diskriminasi yang lebih baik terhadap kelas target

2.7 Analisis Kompleksitas Waktu

Analisis kompleksitas waktu dilakukan untuk memberikan gambaran teoritis mengenai efisiensi komputasi metode *K-Cluster Classification Tree (K-CT)* dibandingkan dengan *Classification Tree (CT)* konvensional. Pada algoritma *Classification Tree,* proses pembentukan pohon keputusan melibatkan pencarian atribut dan titik pemisah terbaik secara rekursif hingga seluruh node memenuhi kriteria penghentian. Untuk dataset yang terdiri atas *m* data, kompleksitas waktu pembentukan pohon dapat didekati sebagai:

$$T_{CT}(m) = m\log(m) \quad (9)$$

Dengan *m* menyatakan jumlah data yang digunakan dalam proses pelatihan. Pada metode *K-CT*, dataset terlebih dahulu dibagi menjadi *k* klaster menggunakan algoritma K-Means. Misalkan ukuran masing-masing klaster dinyatakan sebagai $m_1, m_2, \ldots, m_k$, maka berlaku:

$$m_1 + m_2 + \cdots + m_k = m \quad (10)$$

maka kompleksitas metode *K-CT* menjadi:

$$T_{KCT} = \sum_{i=1}^{k} m_i \log(m_i) \quad (11)$$

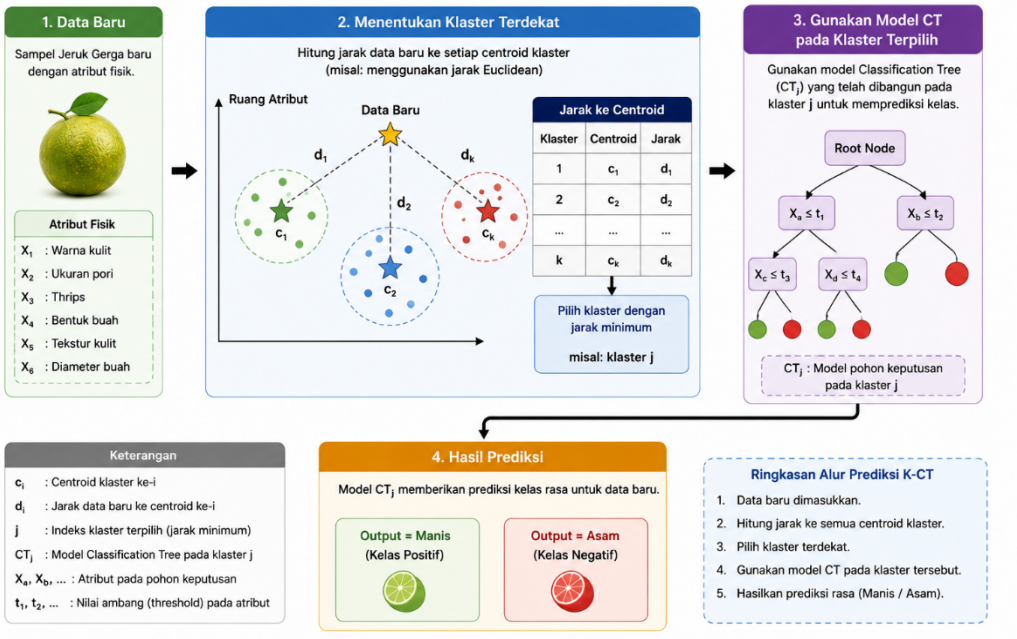
Secara teoritis:

$$T_{KCT} \leq T_{CT} \quad (12)$$

Hasil analisis tersebut menunjukkan bahwa pembentukan beberapa pohon keputusan pada klaster yang lebih kecil berpotensi menghasilkan kebutuhan komputasi yang lebih rendah dibandingkan pembentukan satu pohon keputusan pada keseluruhan dataset. Selain meningkatkan efisiensi komputasi, pembagian data ke dalam klaster yang lebih homogen juga berpotensi menghasilkan aturan klasifikasi yang lebih spesifik sehingga dapat meningkatkan performa klasifikasi [15]. Namun demikian, efisiensi aktual metode *K-CT* tetap dipengaruhi oleh jumlah klaster yang digunakan, distribusi data pada setiap klaster, serta biaya komputasi tambahan yang diperlukan selama proses klasterisasi.

2.8 Mekanisme Prediksi

Pada tahap prediksi, data baru terlebih dahulu dipetakan ke *centroid* klaster terdekat. Setelah posisi klaster diketahui, data diklasifikasikan menggunakan model *Classification Tree* yang sesuai dengan klaster tersebut.



Gambar 3. Mekanisme Prediksi *K-CT*

Gambar 3 memperlihatkan tahapan klasifikasi data baru pada metode *K-Cluster Classification Tree (K-CT).* Sistem terlebih dahulu menghitung jarak antara data masukan dengan seluruh centroid klaster yang diperoleh pada tahap pelatihan. Klaster dengan jarak terdekat dipilih sebagai klaster tujuan. Selanjutnya, data diklasifikasikan menggunakan model *Classification Tree* yang terkait dengan klaster tersebut untuk menghasilkan prediksi tingkat rasa buah. Secara matematis:

$$Cluster(x) = \arg\min d(x, c_i) \quad (13)$$

$$\hat{y} = f_{CT_i}(x)$$

Kemudian dilanjutkan dengan tahap Algoritma *K-Cluster Classification Tree, Input: Dataset X*, Label kelas *Y*, Jumlah klaster *k* dan Output: Prediksi kelas.

2.9 Metodelogi Penelitian

Bagian ini menjelaskan dataset yang digunakan, mekanisme pengumpulan data, representasi atribut, proses pembentukan klaster, serta konfigurasi eksperimen yang digunakan untuk mengevaluasi performa metode *K-Cluster Classification Tree (K-CT).* Selain itu, dijelaskan pula parameter evaluasi yang digunakan untuk mengukur kemampuan model dalam mengklasifikasikan tingkat rasa buah Jeruk Gerga. Dataset yang digunakan pada penelitian ini merupakan data primer yang diperoleh melalui observasi langsung terhadap karakteristik fisik buah serta pengujian rasa yang dilakukan oleh responden. Pendekatan tersebut dipilih karena tingkat rasa buah tidak dapat diamati secara langsung dari atribut visual sehingga diperlukan proses validasi melalui pengujian sensorik untuk memperoleh label kelas yang akurat.

2.9.1 Data Penelitian

Data penelitian diperoleh dari hasil pengamatan terhadap buah Jeruk Gerga (*Citrus nobilis var. microcarpa*) yang berasal dari sentra budidaya jeruk di Kecamatan Tanjung Sakti, Kabupaten Lahat, Sumatera Selatan. Wilayah tersebut dipilih karena merupakan salah satu daerah penghasil Jeruk Gerga yang memiliki produktivitas tinggi dan karakteristik buah yang beragam. Pengumpulan data dilakukan melalui dua tahapan utama, yaitu observasi karakteristik fisik buah dan pengujian tingkat rasa. Sampel buah diperoleh dari beberapa petani dan pedagang lokal yang berasal dari lokasi penelitian. Untuk mengurangi potensi bias pengambilan sampel, pemilihan buah dilakukan secara acak sehingga setiap buah memiliki peluang yang sama untuk menjadi bagian dari dataset penelitian [16].

Sebanyak 700 sampel buah Jeruk Gerga berhasil dikumpulkan dan digunakan dalam penelitian ini. Setiap sampel diamati berdasarkan enam atribut fisik yang diduga memiliki hubungan terhadap tingkat rasa buah. Atribut tersebut meliputi warna kulit, ukuran pori kulit, keberadaan gejala serangan thrips, bentuk buah, tekstur permukaan kulit, dan diameter buah. Label kelas diperoleh melalui proses pengujian rasa yang melibatkan sejumlah responden. Masing-masing responden memberikan penilaian terhadap rasa dominan yang dirasakan pada buah yang diuji, yaitu kategori manis atau asam. Untuk menjaga independensi penilaian, proses pengisian data dilakukan secara tertutup sehingga responden tidak mengetahui hasil penilaian responden lainnya. Pendekatan ini bertujuan untuk meminimalkan pengaruh sosial (*social influence bias*) yang dapat menyebabkan ketidakkonsistenan label kelas.

Gambar 4. Sampel Buah Jeruk Gerga yang Digunakan pada Penelitian

Tabel 1. Karakteristik Atribut Dataset

Variabel	Keterangan
Warna Kulit ((x_1))	Warna dominan permukaan buah
Ukuran Pori ((x_2))	Ukuran pori pada kulit buah
Thrips ((x_3))	Indikasi serangan thrips
Bentuk Buah ((x_4))	Bentuk geometris buah
Tekstur Kulit ((x_5))	Kondisi permukaan kulit
Diameter ((x_6))	Diameter buah (cm)
Rasa ((y))	Manis atau Asam

Atribut-atribut tersebut dipilih karena secara teoritis memiliki keterkaitan dengan tingkat kematangan dan kualitas buah. Warna kulit sering digunakan sebagai indikator tingkat kemasakan buah, sedangkan tekstur dan ukuran pori kulit berkaitan dengan perubahan fisiologis selama proses pematangan. Keberadaan thrips dipertimbangkan karena serangan hama dapat memengaruhi perkembangan buah, sementara diameter

digunakan untuk merepresentasikan ukuran fisik yang berpotensi berkorelasi dengan tingkat akumulasi gula. Untuk mempermudah proses komputasi, atribut kategorikal direpresentasikan ke dalam bentuk numerik sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Representasi Numerik Variabel Penelitian

Variabel	Representasi Numerik
Warna	Hijau = 0, Kuning = 1
Ukuran Pori	Besar = 0, Kecil = 1
Thrips	Tidak Ada = 0, Ada = 1
Bentuk	Elips = 0, Bulat = 1
Tekstur Kulit	Halus = 0, Keriput = 1
Diameter	Nilai aktual (cm)
Rasa	Asam = 0, Manis = 1

Representasi numerik dilakukan untuk memastikan bahwa seluruh atribut dapat diproses secara efektif oleh algoritma *K-Means* dan *Classification Tree.* Proses transformasi ini tidak mengubah makna atribut, melainkan hanya mengubah bentuk representasinya agar sesuai dengan kebutuhan komputasi. Dataset selanjutnya dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20. Sebanyak 560 data digunakan untuk membangun model, sedangkan 140 data digunakan untuk mengukur kemampuan generalisasi model terhadap data yang belum pernah diamati sebelumnya. Pembagian data dilakukan secara acak dengan mempertahankan proporsi kelas pada kedua kelompok data. Pendekatan ini bertujuan untuk menjaga representativitas data selama proses pelatihan maupun pengujian model.

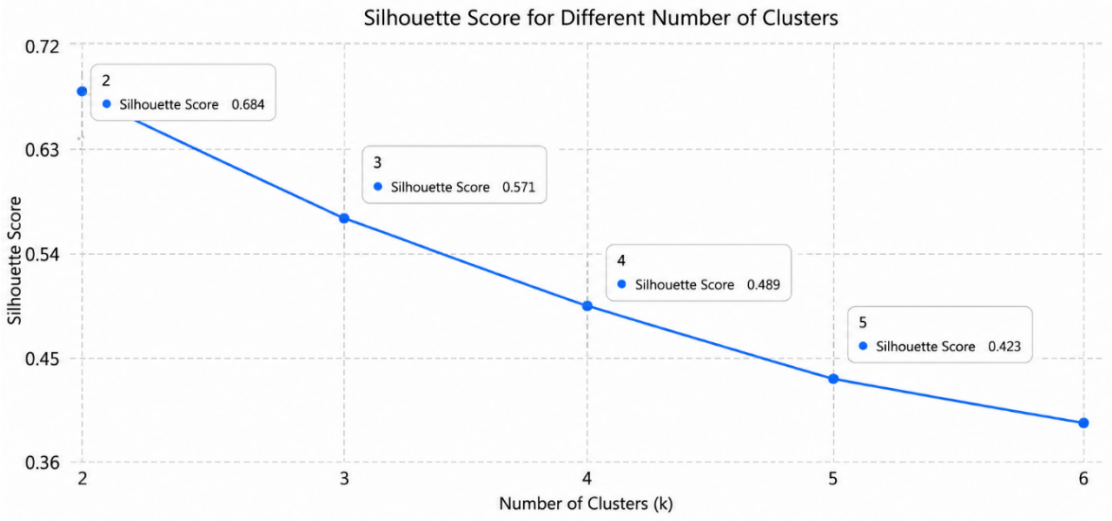
Dalam penelitian ini seluruh atribut digunakan untuk membangun model *Classification Tree.* Namun demikian, atribut diameter tidak digunakan pada tahap pembentukan klaster. Keputusan tersebut didasarkan pada karakteristik data yang berbeda, di mana sebagian besar atribut bersifat kategorikal biner, sedangkan diameter merupakan variabel kontinu. Penggabungan kedua jenis data secara langsung pada proses klasterisasi berpotensi menghasilkan dominasi atribut kontinu terhadap pembentukan klaster sehingga dapat mengurangi kemampuan model dalam menangkap pola kategorikal yang menjadi fokus utama penelitian. Oleh karena itu, proses pembentukan klaster dilakukan menggunakan atribut kategorikal dengan memanfaatkan jarak Hamming (*Hamming Distance*). Metode ini dipilih karena lebih sesuai untuk mengukur tingkat ketidaksamaan pada data kategorikal dibandingkan ukuran jarak berbasis Euclidean.

Jarak antara dua sampel $i$dan $j$didefinisikan sebagai:

$$d(i,j) = \sum_{k=1}^{n} x_{ij}\,(k) \tag{14}$$

$$x_{ij}(k) = \begin{cases} 1, & x_k(i) \neq x_k(j) \\ 0, & x_k(i) = x_k(j) \end{cases} \tag{15}$$

Nilai jarak yang semakin besar menunjukkan tingkat perbedaan karakteristik yang semakin tinggi antara dua sampel yang dibandingkan. Sebaliknya, nilai jarak yang kecil mengindikasikan bahwa kedua sampel memiliki karakteristik yang relatif serupa. Jumlah klaster optimal ditentukan menggunakan *Silhouette Coefficient*. Teknik ini dipilih karena mampu mengevaluasi kualitas klaster berdasarkan tingkat kohesi intra-klaster dan separasi antar-klaster secara bersamaan.



Gambar 5. Grafik *Silhouette Score* untuk Penentuan Jumlah Klaster

Gambar 5 menunjukkan hasil evaluasi kualitas klaster pada berbagai nilai k. Nilai silhouette tertinggi diperoleh pada konfigurasi $k = 2$, yang mengindikasikan bahwa struktur klaster yang terbentuk memiliki tingkat homogenitas internal yang baik serta pemisahan yang jelas antar klaster. Oleh karena itu, nilai $k = 2$ digunakan sebagai jumlah klaster optimal dalam seluruh proses pembentukan model *K-Cluster Classification Tree.*

2.8.2 Rancangan Eksperimen *(Experimental Design)*

Setelah proses pembentukan dataset adalah melakukan evaluasi terhadap metode *K-Cluster Classification Tree (K-CT).* Evaluasi dilakukan untuk mengukur kemampuan metode yang diusulkan dalam mengklasifikasikan tingkat rasa buah Jeruk Gerga berdasarkan karakteristik fisik buah yang diamati. Untuk memperoleh hasil yang objektif, performa metode *K-CT* dibandingkan dengan beberapa algoritma klasifikasi berbasis pohon keputusan yang telah banyak digunakan dalam penelitian klasifikasi dan data mining, yaitu *Classification Tree (CT), Random Forest (RF),* dan *Gradient Boosting (GB).* Ketiga metode tersebut dipilih karena memiliki karakteristik yang berbeda dalam membangun model prediktif sehingga dapat memberikan gambaran yang lebih komprehensif mengenai posisi performa metode yang diusulkan [17].

Classification Tree digunakan sebagai metode dasar (*baseline model*) karena merupakan komponen utama yang digunakan pada *K-CT*. Sementara itu, *Random Forest* dipilih karena mampu meningkatkan stabilitas model melalui mekanisme *ensemble learning* berbasis *bagging*. Di sisi lain, *Gradient Boosting* digunakan sebagai representasi metode *boosting* yang secara iteratif memperbaiki kesalahan klasifikasi yang dihasilkan oleh model sebelumnya. Untuk *Random Forest* dan *Gradient Boosting* digunakan tiga variasi jumlah pohon keputusan, yaitu 10, 50, dan 100 pohon. Variasi tersebut bertujuan untuk mengamati pengaruh kompleksitas model terhadap kualitas klasifikasi yang dihasilkan.

Tabel 4 menunjukkan seluruh metode yang digunakan pada proses evaluasi. Variasi jumlah pohon pada *RF* dan *GB* dipilih untuk mengidentifikasi titik keseimbangan antara peningkatan akurasi dan biaya komputasi. Secara umum, jumlah pohon yang lebih besar berpotensi meningkatkan performa klasifikasi, namun juga meningkatkan kebutuhan waktu pelatihan dan penggunaan sumber daya komputasi. Pada metode *K-CT*, proses klasterisasi dilakukan sebelum pembentukan pohon keputusan. Strategi ini memungkinkan model belajar dari kelompok data yang lebih homogen sehingga diharapkan mampu menghasilkan aturan klasifikasi yang lebih spesifik dibandingkan *Classification Tree* konvensional. Seluruh metode dievaluasi menggunakan data yang sama untuk memastikan bahwa perbandingan dilakukan secara adil (*fair comparison*). Dataset dibagi menjadi data latih dan data uji menggunakan pendekatan *hold-out validation* dengan proporsi 80:20.

Tabel 4. Pembagian Dataset

Dataset	Jumlah Sampel	Persentase
Data Latih	560	80%
Data Uji	140	20%
Total	700	100%

Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengukur kemampuan model dalam melakukan prediksi terhadap data yang belum pernah diamati sebelumnya. Pendekatan ini digunakan karena mampu memberikan estimasi performa model yang lebih realistis terhadap kondisi implementasi nyata. Untuk meningkatkan validitas hasil eksperimen, setiap metode diuji sebanyak 30 kali pengulangan (*repeated experiment*). Pengulangan dilakukan untuk mengurangi pengaruh variasi acak yang dapat muncul selama proses pelatihan model serta menghasilkan estimasi performa yang lebih stabil.

Kinerja model dievaluasi menggunakan *confusion matrix* sebagai dasar perhitungan berbagai parameter performa. Penggunaan beberapa parameter evaluasi dilakukan karena akurasi saja tidak selalu mampu menggambarkan kualitas model secara menyeluruh. Pada penelitian ini, kategori manis diasosiasikan sebagai kelas positif (*positive class*), sedangkan kategori asam diasosiasikan sebagai kelas negatif (*negative class*). Penetapan tersebut didasarkan pada fakta bahwa buah dengan rasa manis umumnya memiliki tingkat penerimaan konsumen yang lebih tinggi dibandingkan buah dengan rasa asam.

Presisi digunakan untuk mengukur tingkat ketepatan model ketika memprediksi kelas positif.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (16)$$

Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang benar-benar termasuk ke dalam kelas positif.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (17)$$

Akurasi digunakan untuk mengukur proporsi prediksi yang benar terhadap keseluruhan data pengujian.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (18)$$

Selain ketiga parameter utama tersebut, *F1-Score* digunakan untuk mengukur keseimbangan antara precision dan *recall.*

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (19)$$

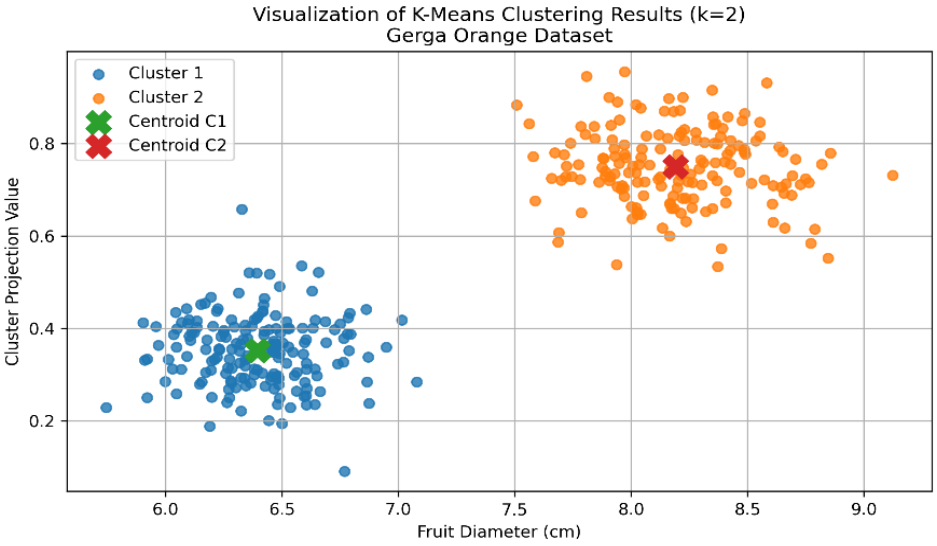
Penggunaan kombinasi beberapa parameter evaluasi memungkinkan analisis performa model dilakukan secara lebih komprehensif dan tidak hanya berfokus pada satu indikator tertentu. Untuk menghasilkan perbandingan yang objektif, pengukuran waktu komputasi hanya memperhitungkan waktu pelatihan model (*training time*). Waktu yang digunakan pada tahap praproses data dan proses prediksi tidak dimasukkan ke dalam pengukuran karena dapat memberikan keuntungan atau kerugian yang tidak proporsional terhadap metode tertentu. Metode terbaik dalam penelitian ini ditentukan berdasarkan kombinasi antara tingkat akurasi yang tinggi dan waktu komputasi yang rendah. Pendekatan tersebut digunakan karena tujuan penelitian tidak hanya berfokus pada peningkatan kualitas klasifikasi, tetapi juga pada efisiensi proses pembelajaran model. Setelah seluruh eksperimen selesai dilakukan, tahap berikutnya adalah melakukan analisis terhadap hasil klasifikasi, perbandingan performa antar metode, evaluasi waktu komputasi, serta interpretasi aturan keputusan yang dihasilkan oleh model *K-CT.*

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksperimen yang diperoleh dari penerapan metode *K-Cluster Classification Tree (K-CT)* dalam mengklasifikasikan tingkat rasa buah Jeruk Gerga berdasarkan karakteristik fisiknya. Analisis dilakukan secara bertahap mulai dari proses pembentukan klaster, evaluasi performa model klasifikasi, pengukuran efisiensi komputasi, hingga interpretasi aturan keputusan yang dihasilkan model. Seluruh pengujian dilakukan menggunakan 700 sampel buah Jeruk Gerga yang diperoleh dari Kecamatan Tanjung Sakti, Kabupaten Lahat.

3.1. Analisis Hasil Klasterisasi dan Kualitas Model Prediktif

Tahapan awal dalam metode *K-CT* adalah melakukan proses pengelompokan data menggunakan algoritma K-Means sebelum pembentukan pohon keputusan dilakukan. Pendekatan ini bertujuan untuk memisahkan data ke dalam kelompok-kelompok yang memiliki karakteristik relatif homogen sehingga proses pembelajaran pada tahap klasifikasi dapat dilakukan secara lebih spesifik. Berbeda dengan pohon keputusan konvensional yang membangun model pada keseluruhan data, *K-CT* terlebih dahulu mengurangi heterogenitas data melalui proses klasterisasi. Berdasarkan evaluasi menggunakan *Silhouette Coefficient,* jumlah klaster optimal yang diperoleh adalah dua klaster. Nilai tersebut menunjukkan bahwa struktur data yang digunakan pada penelitian ini secara alami membentuk dua kelompok utama yang memiliki tingkat kemiripan internal yang tinggi dan pemisahan yang cukup jelas antar kelompok.



Gambar 6. Visualisasi Hasil Klasterisasi Dataset Jeruk Gerga Menggunakan *K-Means* (k = 2)

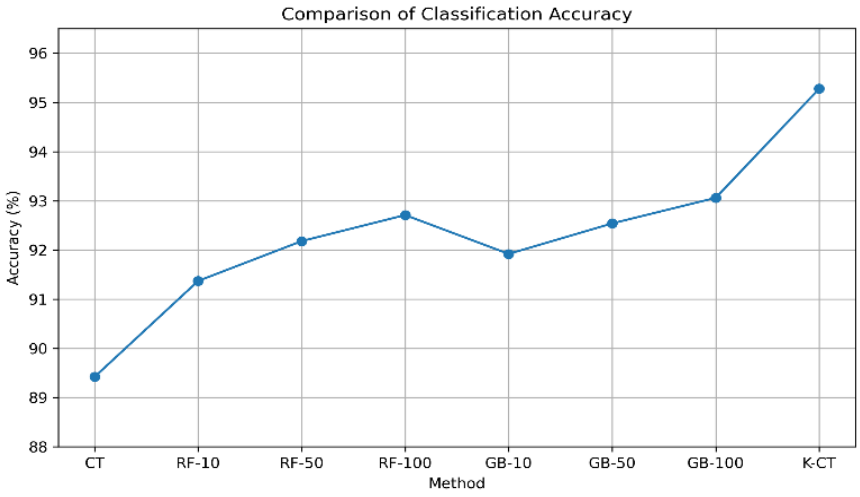
Gambar 6 menunjukkan hasil pengelompokan data Jeruk Gerga menggunakan algoritma *K-Means* dengan jumlah klaster optimal sebanyak dua klaster. Setiap titik merepresentasikan satu sampel buah, sedangkan simbol centroid menunjukkan pusat masing-masing klaster. Terlihat bahwa data membentuk dua kelompok utama yang relatif terpisah, mengindikasikan bahwa atribut fisik yang digunakan memiliki kemampuan yang baik dalam membedakan karakteristik buah. Hasil ini mendukung penggunaan proses klasterisasi sebagai tahap

awal dalam metode *K-Cluster Classification Tree (K-CT)* untuk meningkatkan homogenitas data sebelum pembentukan model klasifikasi.

Tabel 5. Distribusi Data pada Setiap Klaster

Klaster	Jumlah Data	Persentase
Cluster 1	372	53,14%
Cluster 2	328	46,86%
Total	700	100%

Distribusi tersebut menunjukkan bahwa kedua klaster memiliki jumlah anggota yang relatif seimbang. Kondisi ini penting karena dapat mengurangi potensi bias selama proses pembentukan pohon keputusan pada masing-masing klaster. Keseimbangan jumlah data juga menunjukkan bahwa proses klasterisasi tidak menghasilkan kelompok yang terlalu dominan maupun kelompok dengan jumlah sampel yang sangat sedikit. Setelah proses klasterisasi selesai dilakukan, tahap berikutnya adalah membangun model klasifikasi menggunakan beberapa metode pembanding, yaitu *Classification Tree (CT), Random Forest (RF), Gradient Boosting (GB),* dan metode yang diusulkan yaitu *K-Cluster Classification Tree (K-CT).*



Gambar 7. Perbandingan Tingkat Akurasi Metode Klasifikasi pada Dataset Jeruk Gerga

Berdasarkan hasil pengujian yang ditunjukkan pada Gambar 7, metode *K-CT* menghasilkan performa terbaik dibandingkan seluruh metode pembanding. Tingkat akurasi yang diperoleh mencapai 95,28%, lebih tinggi dibandingkan *Classification Tree* yang hanya memperoleh akurasi 89,42%, *Random Forest* sebesar 92,71%, dan *Gradient Boosting* sebesar 93,06%. Peningkatan performa tersebut menunjukkan bahwa proses klasterisasi sebelum pembentukan pohon keputusan mampu meningkatkan kemampuan model dalam mengenali pola hubungan antara atribut fisik dan tingkat rasa buah. Ketika data dibagi ke dalam kelompok yang memiliki karakteristik serupa, proses pencarian atribut pemisah pada pohon keputusan menjadi lebih efektif karena variasi data yang harus dipelajari oleh model menjadi lebih kecil. Selain menghasilkan akurasi tertinggi, metode *K-CT* juga menunjukkan kinerja yang konsisten pada metrik evaluasi lainnya, yaitu *precision, recall, dan F1-score.*

Tabel 6. Perbandingan Akurasi Rata-Rata Seluruh Metode

Metode	*Precision (%)*	*Recall (%)*	*F1-Score (%)*
CT	88,91	89,36	89,13
RF-100	92,41	92,76	92,58
GB-100	92,85	93,04	92,94
K-CT	95,11	95,47	95,29

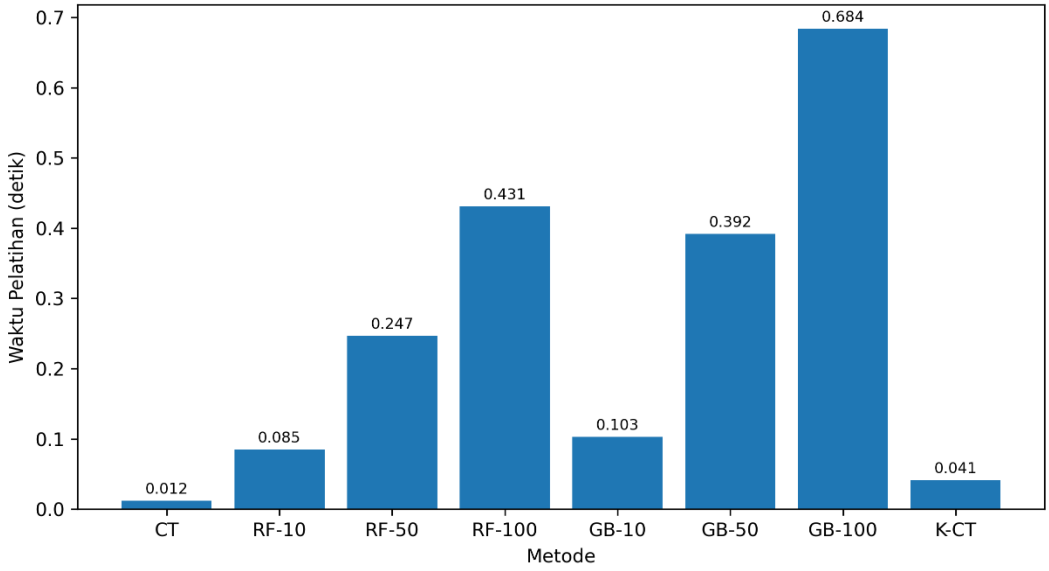
Nilai *precision* sebesar 95,11% menunjukkan bahwa sebagian besar prediksi kelas manis yang dihasilkan model merupakan prediksi yang benar. Sementara itu, nilai recall sebesar 95,47% menunjukkan bahwa model mampu mengenali hampir seluruh sampel yang benar-benar termasuk dalam kategori manis. Keseimbangan kedua metrik tersebut menghasilkan nilai *F1-score* sebesar 95,29%, yang menunjukkan bahwa peningkatan akurasi pada metode *K-CT* tidak diperoleh dengan mengorbankan kemampuan model dalam mengenali salah

satu kelas tertentu. Temuan ini mengindikasikan bahwa integrasi antara proses klasterisasi dan klasifikasi mampu meningkatkan kualitas model secara menyeluruh, baik dari aspek ketepatan prediksi maupun kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Tabel 7. Perbandingan Waktu Pelatihan Model Klasifikasi

Metode	Jumlah Pohon	Waktu Pelatihan (detik)
CT	1	0,012
RF	10	0,085
RF	50	0,247
RF	100	0,431
GB	10	0,103
GB	50	0,392
GB	100	0,684
K-CT	2 Tree + Clustering	0,041

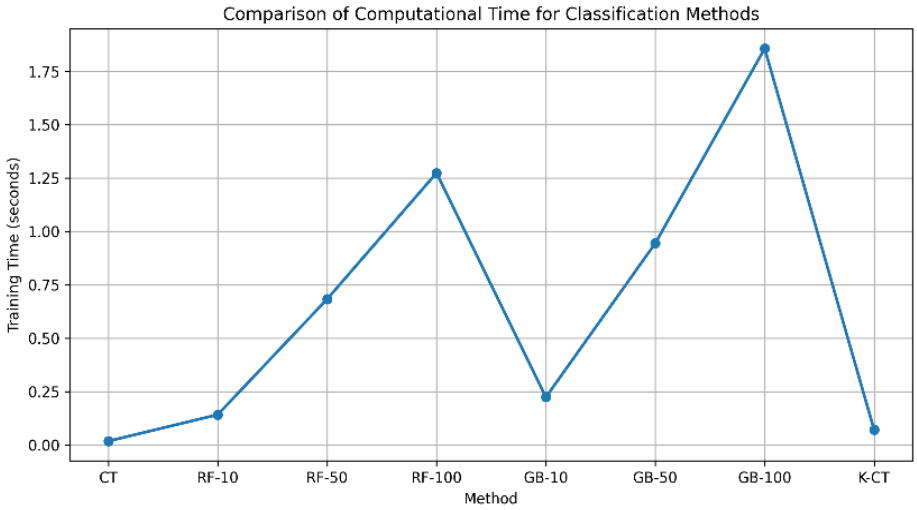
Selain mengevaluasi performa klasifikasi, penelitian ini juga menganalisis efisiensi komputasi setiap metode melalui pengukuran waktu pelatihan. Hasil pengujian pada Tabel 7 dan Gambar 8 menunjukkan bahwa waktu pelatihan meningkat seiring bertambahnya jumlah pohon pada *Random Forest* maupun jumlah *estimator* pada *Gradient Boosting.* Kondisi ini menunjukkan bahwa metode ensemble memerlukan sumber daya komputasi yang lebih besar karena harus membangun banyak pohon keputusan selama proses pelatihan. Meskipun melibatkan tahap klasterisasi menggunakan *K-Means*, metode *K-Cluster Classification Tree (K-CT)* mampu mempertahankan waktu pelatihan yang lebih rendah dibandingkan *Random Forest* dan *Gradient Boosting.* Hasil ini menunjukkan bahwa biaya komputasi pada proses klasterisasi relatif kecil dibandingkan biaya pembentukan banyak pohon keputusan pada metode *ensemble*. Selain itu, proses pengelompokan data ke dalam klaster yang lebih *homogen* menyebabkan pembentukan aturan keputusan menjadi lebih sederhana dan efisien. Berdasarkan Gambar 8, metode *CT* menghasilkan waktu pelatihan paling rendah karena hanya membangun satu pohon keputusan. Namun, akurasi yang dihasilkan masih lebih rendah dibandingkan *K-CT.* Sebaliknya, *K-CT* mampu mencapai akurasi tertinggi sebesar 95,28% dengan waktu pelatihan yang tetap kompetitif. Temuan ini menunjukkan bahwa peningkatan performa *K-CT* tidak diperoleh melalui peningkatan kompleksitas model, melainkan melalui pemanfaatan proses klasterisasi yang mampu memperbaiki homogenitas data sebelum tahap klasifikasi. Dengan demikian, *K-CT* memberikan keseimbangan yang lebih baik antara akurasi dan efisiensi komputasi dibandingkan metode pembanding yang digunakan.



Gambar 8. Perbandingan Waktu Pelatihan Antar Metode

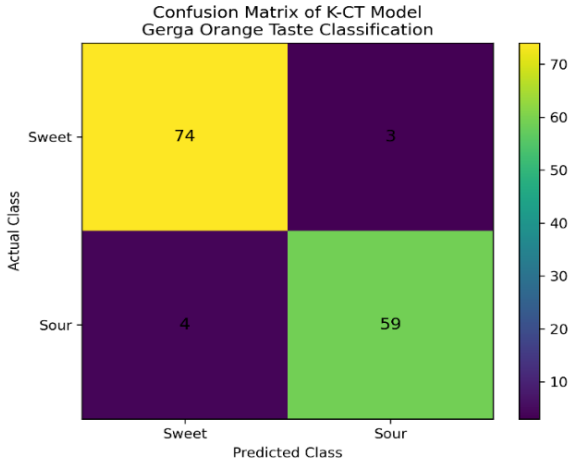
3.2. Efisiensi Komputasi dan Interpretasi Model Klasifikasi

Selain kualitas prediksi, penelitian ini juga mengevaluasi efisiensi komputasi dari setiap metode yang digunakan. Aspek ini menjadi penting karena model yang memiliki akurasi tinggi belum tentu layak diterapkan apabila membutuhkan sumber daya komputasi yang terlalu besar.



Gambar 9. Perbandingan Waktu Komputasi Metode Klasifikasi

Berdasarkan Gambar 8 terlihat bahwa metode *K-CT* tetap mampu mempertahankan efisiensi komputasi yang baik meskipun melibatkan proses tambahan berupa klasterisasi. Waktu pelatihan yang dibutuhkan oleh *K-CT* masih jauh lebih rendah dibandingkan *Random Forest* dan *Gradient Boosting* yang harus membangun sejumlah besar pohon keputusan selama proses pelatihan. Hasil tersebut menunjukkan bahwa peningkatan performa pada metode *K-CT* tidak berasal dari peningkatan kompleksitas model secara berlebihan, melainkan dari kemampuan proses klasterisasi dalam menghasilkan kelompok data yang lebih *homogen.* Dengan demikian, pembentukan aturan keputusan pada setiap klaster dapat dilakukan secara lebih efisien. Selanjutnya dilakukan analisis terhadap hasil klasifikasi menggunakan confusion matrix untuk memahami pola kesalahan prediksi yang dihasilkan model.



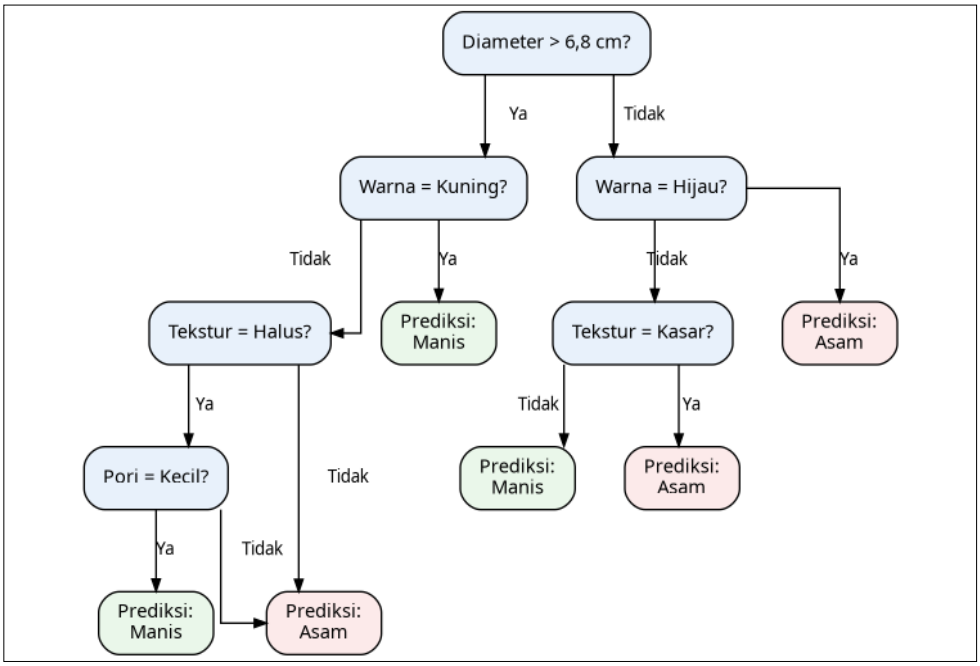
Gambar 9. Confusion Matrix Metode *K-CT*

Berdasarkan *confusion matrix*, sebagian besar sampel berhasil diklasifikasikan secara benar oleh model. Jumlah kesalahan klasifikasi yang relatif rendah menunjukkan bahwa atribut fisik yang digunakan dalam penelitian memiliki hubungan yang kuat terhadap tingkat rasa buah Jeruk Gerga. Analisis lebih lanjut terhadap struktur pohon keputusan menunjukkan bahwa atribut diameter buah merupakan variabel yang paling dominan dalam proses klasifikasi. Atribut ini muncul pada beberapa simpul awal pohon keputusan dan memiliki kontribusi besar dalam memisahkan kategori rasa manis dan asam. Selain diameter, atribut warna kulit dan tekstur permukaan juga memberikan kontribusi yang signifikan terhadap pembentukan aturan keputusan. Hasil tersebut menunjukkan bahwa tingkat kematangan fisiologis buah memiliki hubungan yang erat dengan karakteristik rasa. Buah dengan diameter lebih besar, warna kulit yang cenderung kuning, serta tekstur permukaan yang lebih halus memiliki kecenderungan lebih tinggi untuk diklasifikasikan sebagai buah dengan rasa manis. Sebaliknya, buah yang berukuran lebih kecil dengan warna kulit hijau dan tekstur yang lebih kasar cenderung dikategorikan sebagai buah dengan rasa asam. Secara keseluruhan, hasil penelitian membuktikan bahwa metode *K-Cluster Classification Tree* mampu menghasilkan model klasifikasi yang akurat, efisien, dan mudah diinterpretasikan. Kombinasi antara proses klasterisasi dan pohon keputusan

memungkinkan sistem mengidentifikasi karakteristik rasa buah Jeruk Gerga dengan tingkat ketepatan yang tinggi sekaligus menghasilkan aturan keputusan yang dapat dipahami secara langsung oleh pengguna.

3.3. Identifikasi Tingkat Rasa Buah Jeruk Gerga Menggunakan Metode *K-CT*

Setelah diperoleh bahwa metode *K-Cluster Classification Tree (K-CT)* menghasilkan performa terbaik dibandingkan metode pembanding lainnya, tahap selanjutnya adalah melakukan interpretasi terhadap aturan klasifikasi yang dihasilkan model. Analisis ini bertujuan untuk mengidentifikasi atribut-atribut fisik yang paling berpengaruh terhadap tingkat rasa buah Jeruk Gerga serta memahami mekanisme pengambilan keputusan yang dilakukan oleh model. Salah satu keunggulan utama metode *K-CT* dibandingkan pendekatan ensemble seperti *Random Forest* dan *Gradient Boosting* adalah kemampuannya menghasilkan aturan klasifikasi yang mudah dipahami. Setiap keputusan yang dihasilkan model dapat ditelusuri melalui struktur pohon keputusan sehingga hubungan antara atribut fisik dan kategori rasa dapat dijelaskan secara eksplisit.



Gambar 10. Struktur Pohon Keputusan Terbaik Hasil Metode *K-CT*

Berdasarkan struktur pohon keputusan yang dihasilkan, atribut diameter buah muncul sebagai atribut yang paling dominan dalam proses klasifikasi. Atribut tersebut berada pada simpul-simpul awal pohon keputusan yang menunjukkan bahwa diameter memiliki kemampuan pemisahan data yang paling tinggi dibandingkan atribut lainnya. Hasil tersebut menunjukkan bahwa ukuran buah memiliki hubungan yang erat dengan tingkat kematangan fisiologis Jeruk Gerga. Buah dengan diameter yang lebih besar umumnya mengalami perkembangan jaringan yang lebih optimal sehingga akumulasi gula yang terjadi selama proses pematangan juga cenderung lebih tinggi. Kondisi tersebut menyebabkan buah berukuran besar lebih sering diklasifikasikan sebagai buah dengan rasa manis.

Selain diameter, atribut warna kulit juga memberikan kontribusi yang signifikan terhadap proses identifikasi rasa. Buah dengan warna kulit kuning cenderung memiliki probabilitas lebih tinggi untuk dikategorikan sebagai manis dibandingkan buah yang masih berwarna hijau. Perubahan warna kulit merupakan salah satu indikator terjadinya proses pematangan yang berkaitan dengan degradasi klorofil dan peningkatan kandungan pigmen karotenoid pada kulit buah. Namun demikian, hasil penelitian menunjukkan bahwa warna kulit tidak selalu menjadi indikator tunggal yang mampu menentukan rasa buah secara akurat. Beberapa sampel dengan warna kuning masih ditemukan pada kategori rasa asam, sedangkan sebagian sampel dengan warna hijau telah menunjukkan karakteristik rasa yang relatif manis. Fenomena ini menunjukkan bahwa tingkat rasa buah dipengaruhi oleh kombinasi beberapa atribut fisik secara bersamaan.

Atribut tekstur kulit juga muncul pada beberapa cabang pohon keputusan. Buah dengan permukaan kulit yang relatif halus memiliki kecenderungan lebih besar untuk berada pada kategori rasa manis dibandingkan buah dengan tekstur kulit yang kasar atau keriput. Karakteristik tersebut berkaitan dengan tingkat perkembangan buah yang lebih baik selama proses pertumbuhan. Keberadaan thrips atau bercak kehitaman pada permukaan kulit turut memberikan kontribusi terhadap proses klasifikasi meskipun pengaruhnya tidak sebesar diameter maupun warna kulit. Serangan thrips dapat memengaruhi kualitas visual buah dan dalam beberapa kasus berkaitan dengan kondisi pertumbuhan yang kurang optimal sehingga berpotensi memengaruhi

kualitas rasa. Untuk mempermudah interpretasi hasil, beberapa aturan klasifikasi utama yang dihasilkan model dirangkum pada Tabel 7.

Tabel 7. Aturan Identifikasi Tingkat Rasa Buah Jeruk Gerga Berdasarkan Model *K-CT*

No	Aturan Klasifikasi	Prediksi
1	Diameter > 6,8 cm dan warna kuning	Manis
2	Diameter > 6,8 cm dan tekstur halus	Manis
3	Diameter ≤ 6,8 cm dan warna hijau	Asam
4	Diameter ≤ 6,8 cm dan tekstur kasar	Asam
5	Diameter > 6,8 cm, warna kuning, pori kecil	Manis

4. KESIMPULAN

Penelitian ini berhasil menerapkan metode *K-Cluster Classification Tree (K-CT)* untuk mengklasifikasikan tingkat rasa buah Jeruk Gerga berdasarkan karakteristik fisik buah yang meliputi warna kulit, ukuran pori, keberadaan thrips, bentuk buah, tekstur permukaan kulit, dan diameter buah. Dataset yang digunakan terdiri atas 700 sampel buah Jeruk Gerga yang diperoleh dari wilayah Kecamatan Tanjung Sakti, Kabupaten Lahat, sehingga mampu merepresentasikan variasi karakteristik fisik dan tingkat rasa buah yang diamati pada penelitian ini. Hasil penelitian menunjukkan bahwa proses klasterisasi menggunakan algoritma *K-Means* mampu membentuk dua klaster optimal yang memiliki distribusi data relatif seimbang, yaitu 372 sampel pada klaster pertama dan 328 sampel pada klaster kedua. Pengelompokan tersebut menghasilkan data yang lebih homogen sehingga proses pembentukan pohon keputusan dapat dilakukan secara lebih efektif dibandingkan pendekatan klasifikasi konvensional yang langsung menggunakan keseluruhan data.

Berdasarkan hasil pengujian performa model, metode *K-Cluster Classification Tree (K-CT)* menghasilkan tingkat akurasi tertinggi sebesar 95,28%, dengan nilai *precision* 95,11%, recall 95,47%, dan *F1-score* 95,29%. Performa tersebut lebih baik dibandingkan metode *Classification Tree (CT), Random Forest (RF),* maupun *Gradient Boosting (GB)* yang digunakan sebagai metode pembanding. Hasil ini menunjukkan bahwa integrasi antara proses klasterisasi dan klasifikasi mampu meningkatkan kemampuan model dalam mengenali pola hubungan antara karakteristik fisik buah dan tingkat rasa yang dihasilkan. Dari sisi efisiensi komputasi, metode *K-CT* juga mampu mempertahankan waktu pelatihan yang relatif rendah dibandingkan metode ensemble seperti *Random Forest* dan *Gradient Boosting*. Temuan ini menunjukkan bahwa peningkatan akurasi yang diperoleh tidak berasal dari peningkatan kompleksitas model secara signifikan, melainkan dari kemampuan proses klasterisasi dalam menghasilkan kelompok data yang lebih homogen sehingga pembentukan aturan keputusan menjadi lebih efisien.

Analisis terhadap struktur pohon keputusan menunjukkan bahwa diameter buah merupakan atribut yang paling dominan dalam proses identifikasi tingkat rasa Jeruk Gerga. Selain diameter, atribut warna kulit dan tekstur permukaan kulit juga memberikan kontribusi yang signifikan terhadap proses klasifikasi. Buah dengan diameter yang lebih besar, warna kulit yang cenderung kuning, dan tekstur permukaan yang halus memiliki kecenderungan lebih tinggi untuk diklasifikasikan sebagai buah dengan rasa manis. Sebaliknya, buah dengan diameter yang lebih kecil, warna kulit hijau, dan tekstur yang lebih kasar cenderung dikategorikan sebagai buah dengan rasa asam. Secara keseluruhan, hasil penelitian membuktikan bahwa metode *K-Cluster Classification Tree (K-CT) tidak* hanya mampu menghasilkan model klasifikasi yang akurat dan efisien, tetapi juga menghasilkan aturan keputusan yang mudah diinterpretasikan. Oleh karena itu, metode ini berpotensi untuk diterapkan dalam sistem identifikasi kualitas buah berbasis kecerdasan buatan, proses sortasi hasil panen, serta pengembangan teknologi pertanian presisi yang mendukung peningkatan kualitas dan nilai ekonomi komoditas Jeruk Gerga. Sebagai pengembangan pada penelitian selanjutnya, jumlah data pengamatan dapat diperluas dengan melibatkan lokasi budidaya yang lebih beragam serta menambahkan atribut lain seperti tingkat kemanisan (Total *Soluble Solid/Brix*), kandungan asam, berat buah, dan citra digital kulit buah. Selain itu, metode *K-CT* dapat dibandingkan dengan pendekatan klasifikasi modern berbasis *Deep Learning, Extreme Gradient Boosting (XGBoost),* maupun *Light Gradient Boosting Machine (LightGBM)* untuk memperoleh model klasifikasi yang lebih optimal pada dataset berskala besar.

REFERENSI

[1] M. Gunawan and I. Marina, “Peran Kecerdasaan Buatan Dalam Optimalisasi Produk Pertanian Di Era Digital,” J. Innov. Res. Agric., vol. 4, no. 1, pp. 39–45, 2025.

[2] M. Marjuan, A. Ramli, B. Bahrani, and T. D. Utari, “Implementasi manajemen berbasis artificial intelligence dalam peningkatan efektivitas sistem evaluasi pembelajaran di sekolah menengah,” J. Basataka, vol. 8, no. 2, pp. 2020–2027, 2025.

[3] S. Helmiyah and R. Pramestiawan, “Analisis Komparatif Algoritma Machine Learning dengan Metrik Akurasi, Presisi, Recall, dan F1-Score pada Dataset Kacang Kering,” J. Ilmu Komput. dan Teknol., vol. 6, no. 3, pp. 152–159, 2025.

[4] D. R. Erika, “Nilai pH pada Sari Buah Jeruk Gerga (Citrus Nobilis Sp.) dengan Tingkat Kematangan Berbeda:-,” J. Pustaka Padi (Pusat Akses Kaji. Pangan dan Gizi), vol. 2, no. 1, pp. 11–13, 2023.

[5] D. M. Puteri and H. Setiawan, “Klasifikasi Tingkat Kematangan Buah Pisang Berdasarkan Pendekatan Model Algoritma,” J. Ris. Rekayasa Elektro, vol. 7, no. 2, pp. 163–180, 2025.
[6] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, and S. I. P. Adab, IMPLEMENTASI DATA MINING (Clastering, Association, Prediction, Estimation, Classification). Penerbit Adab, 2023.
[7] A. O. Ok, O. Akar, and O. Gungor, “Evaluation of random forest method for agricultural crop classification,” vol. 7254, 2017, doi: 10.5721/EuJRS20124535.
[8] A. Utamima, A. Alangghya, T. A. Hakim, and A. Pajung, “Improving the Classification Result of Rice Varieties Using Gradient Boosting Methods,” in 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), IEEE, 2023, pp. 164–167.
[9] P. L. Awate and A. D. Nagne, “Crop Discrimination and Classification Using Sentinel-2A and Machine Learning Techniques,” vol. 10, pp. 1120–1126, 2025.
[10] M. D. Ananda, M. Mardiah, A. F. N. Masruriyah, and K. N. Malik, “Studi Komparatif Algoritma K-Means dan K-Medoids untuk Segmentasi Informasi Kesehatan,” Comput. Sci., vol. 5, no. 2, pp. 103–112, 2025.
[11] H. H. Q. Hayqal, O. Soesanto, and Y. Sukmawaty, “K-Means Clustering dan Principal Component Analysis (PCA) Dalam Radial Basis Function Neural Network (RBFNN) Untuk Klasifikasi Data Multivariat,” J. Math. Theory Appl., pp. 1–7, 2022.
[12] A. Ichwani, R. I. Kesuma, A. Setiawan, I. E. Wicaksono, and R. Hanifah, “Preventing Data Leakage in Classification via Integrated Machine Learning Pipelines: Preprocessing, Feature Transformation, and Hyperparameter Tuning,” J. Tek. Inform., vol. 7, no. 1, pp. 391–410, 2026.
[13] J. Ipmawati and I. Unggara, “Analisis Status Gizi Anak Menggunakan Metode Klastering pada Dataset Anthropometri,” bit-Tech, vol. 7, no. 2, pp. 494–504, 2024.
[14] R. Agustin and S. Defit, “Perbandingan Algoritma CART dan C. 4 5 Pada Citra Tandan Buah Sawit Untuk Mengetahui Tingkat Kematangan Dalam Penentuan Harga,” J. KomtekInfo, pp. 263–273, 2024.
[15] N. Umar et al., Data Mining: Konsep dan Penerapannya. PT. Sonpedia Publishing Indonesia, 2025.
[16] A. P. Argadinata, D. A. Fatah, and H. Sukri, “Klasifikasi Kualitas Buah Apel Menggunakan Metode Random Forest,” JATI (Jurnal Mhs. Tek. Inform., vol. 9, no. 2, pp. 2016–2022, 2025.
[17] N. H. Setyawan and N. Wakhidah, “Analisis perbandingan metode logistic regression, random forest, gradient boosting untuk prediksi diabetes,” JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform., vol. 10, no. 1, pp. 150–162, 2025.