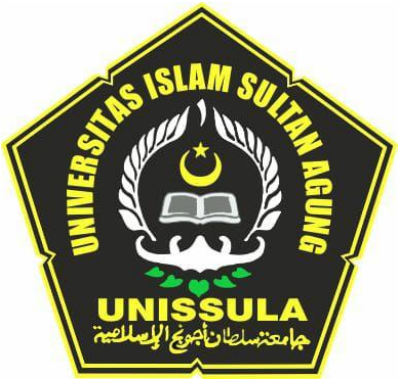
PENERAPAN ALGORITMA *K-MEANS CLUSTERING* UNTUK SEGMENTASI PENYAKIT DAUN MANGGA

LAPORAN TUGAS AKHIR

Laporan ini disusun untuk memenuhi salah satu syarat memperoleh
Gelar Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika
Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang



DISUSUN OLEH:

ISTAKNAFA ASCETIC NAYA
32602100061

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG
2025

FINAL PROJECT

PENERAPAN ALGORITMA *K-MEANS CLUSTERING* UNTUK SEGMENTASI PENYAKIT DAUN MANGGA

Proposed to complete the requirement to obtain a bachelor's degree (S1)
at Informatics Engineering Departement of Industrial Technology Faculty
Sultan Agung Islamic University

Arranged By:

ISTAKNAFA ASCETIC NAYA
32602100061

MAJORING OF INFORMATICS ENGINEERING
FACULTY OF INDUSTRIAL TECHNOLOGY
SULTAN AGUNG ISLAMIC UNIVERSITY
SEMARANG

2025

LEMBAR PENGESAHAN
TUGAS AKHIR

PENERAPAN ALGORITMA *K-MEANS CLUSTERING* UNTUK SEGMENTASI PENYAKIT DAUN MANGGA

ISTAKNAFA ASCETIC NAYA
NIM 32602100061

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 15 Mei 2025

TIM PENGUJI UJIAN SARJANA :

Moch Taufik, ST, MIT NIDN. 0622037502 (Ketua Penguji)	.....	04/06/2025
Andi Riansyah, ST, M.Kom NIDN. 0609108802 (Anggota Penguji)	.....	04/06/2025
Bagus S.W.P., S.Kom, M.Cs NIDN. 1027118801 (Pembimbing)	.....	04/06/2025

Semarang, 04 Juni 2025
Mengetahui,
Kaprodi Teknik Informatika
Universitas Islam Sultan Agung

Moch. Taufik, ST., MIT
NIDN. 0622037502

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Istaknafa Ascetic Naya

NIM : 32602100061

Judul Tugas Akhir : Penerapan Algoritma *K-Means Clustering* Untuk Segmentasi Penyakit Daun Mangga

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 04 Juni 2025

Yang Menyatakan

Istaknafa Ascetic Naya

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Istaknafa Ascetic Naya

NIM : 32602100061

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul :
Penerapan Algoritma *K-Means Clustering* Untuk Segmentasi Penyakit Daun Mangga

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan Agung.

Semarang, 04 Juni 2025

Yang Menyatakan

Istaknafa Ascetic Naya

KATA PENGANTAR

Dengan mengucapkan rasa syukur alhamdulillah atas kehadirat Allah SWT yang telah memberikan Rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Laporan Tugas Akhir dengan judul "Penerapan Algoritma *K-Means Clustering* untuk Segmentasi Penyakit Daun Mangga" ini untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S-1) pada program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung.

Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada:

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, S.H., M.H yang mengizinkan penulis menimba ilmu dikampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Ir. Novi Marlyana, S.T., M.T., IPU., ASEAN. Eng.
3. Dosen Pembimbing penulis Bapak Bagus Satrio Waluyo Poetro S.Kom., M.Cs yang tulus dan ikhlas memberikan bimbingan dan saran yang berharga kepada penulis selama menyusun laporan ini.
4. Orang tua dan keluarga penulis yang menjadi *support system* dan mengizinkan untuk menyelesaikan laporan ini.
5. Para sahabat, teman-teman yang telah memberikan begitu banyak bantuan, semangat, inspirasi, pengambilan data dan diskusi progres penyusunan Tugas Akhir.

Semarang, 15 Mei 2025

Istaknafa Ascetic Naya

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

ABSTRAK

Penyakit pada daun mangga *(Mangifera indica)* adalah salah satu tantangan utama dalam pertanian yang dapat menyebabkan penurunan kualitas dan kuantitas hasil panen. Penelitian ini bertujuan untuk menerapkan algoritma *K-Means* dalam segmentasi citra daun mango yang terinfeksi penyakit, dengan harapan dapat meningkatkan akurasi deteksi serta memberikan solusi efisien bagi petani. Metodologi yang diterapkan dalam penelitian ini melibatkan pengumpulan citra daun mangga yang sehat dan terinfeksi, *preprocessing* citra untuk meningkatkan kualitas gambar, serta penerapan algoritma *K-Means* untuk mengelompokkan piksel berdasarkan kesamaan warna dan tekstur. Hasil segmentasi kemudian dievaluasi menggunakan metrik akurasi, presisi, dan recall untuk menilai kinerja algoritma dalam memisahkan area sehat dan terinfeksi. Hasil penelitian menunjukkan bahwa algoritma *K-Means* dapat dengan efektif mengidentifikasi dan mengklasifikasikan penyakit pada daun mangga, dengan tingkat akurasi yang memuaskan. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pemantauan kesehatan tanaman berbasis citra, serta membantu petani dalam mengambil langkah pencegahan yang tepat terhadap penyakit tanaman. Dengan penerapan teknologi ini, diharapkan produktivitas dan keberlanjutan dalam sektor pertanian dapat meningkat.

Kata kunci: Algoritma K-Means, Segmentasi Citra, Penyakit Daun Mangga, Pengolahan Citra Digital.

ABSTRACT

Mango leaf disease (Mangifera indica) is one of the main challenges in agriculture that can result in decreased quality and quantity of harvest. This study aims to apply the K-Means algorithm in image segmentation of mango leaves infected with disease, with the hope of increasing detection accuracy and providing efficient solutions for farmers. The methodology used in this study includes collecting images of healthy and infected mango leaves, image preprocessing to improve image quality, and applying the K-Means algorithm to group pixels based on color and texture similarities. The segmentation results are then evaluated using accuracy, precision, and recall metrics to assess the performance of the algorithm in separating healthy and infected areas. The results show that the K-Means algorithm can effectively identify and classify diseases in mango leaves, with a satisfactory level of accuracy. This study is expected to contribute to the development of an image-based plant health monitoring system, as well as assist farmers in taking appropriate preventive measures against plant diseases. With the application of this technology, it is hoped that productivity and sustainability in the agricultural sector can be increased.

Keywords: Algorithm K-Means, Image Segmentation, Mango Leaf Disease, Image Processing

BAB I
PENDAHULUAN

1.1 Latar Belakang

Tanaman mangga (*Mangifera Indica L)* memiliki kemungkinan terserang penyakit. Beberapa jenis penyakit yang umumnya tanaman mangga adalah lalat empedu, *Antracnose*, jamur jelaga dan lainnya (Rosadi dkk, 2023). Adanya penyakit pada tanaman mangga dapat menghambat produksi dan pertumbuhan mangga di Indonesia sehingga kesehatan tanaman ini sangat berpengaruh terhadap perekonomian petani (Hidayat, 2022).

Kemajuan besar dalam teknologi pertanian termasuk otomatisasi proses pengendalian mutu dan pengelolaan tanaman untuk meningkatkan produktivitas dan keamanan pangan (Poetro dkk., 2024). Teknologi pengolahan citra digital menawarkan solusi untuk mengidentifikasi penyakit tanaman secara otomatis (Solikin, 2020). Segmentasi yang bertujuan untuk membedakan objek dari latar belakang gambar adalah teknik pengolahan gambar yang umum digunakan. Petani dapat mengidentifikasi dan mengklasifikasikan penyakit pada daun mangga dengan lebih mudah dengan segmentasi yang efektif (Kumar & Singh, 2019).

Dalam pengolahan citra, algoritma *K-Means* adalah salah satu teknik segmentasi yang paling populer dan efektif. Metode ini mengelompokkan piksel citra berdasarkan kesamaan fitur, seperti warna dan tekstur, yang sangat penting untuk mendeteksi penyakit pada daun. Penerapan Algoritma *K-Means* akan meningkatkan akurasi deteksi penyakit dan memberikan informasi yang bermanfaat bagi petani (Hidayat, 2022).

Oleh karena itu, tujuan penelitian ini adalah untuk menggunakan algoritma *K-Means* untuk segmentasi penyakit daun mangga dan mengevaluasi seberapa efektif untuk menemukan dan mengklasifikasikan penyakit. Penelitian ini diharapkan dapat berkontribusi pada pengembangan teknologi pertanian yang lebih efisien dan efektif.

1.2 Perumusan Masalah

Bagaimana sistem menggunakan algoritma *K-Means Clustering* untuk membedakan dan mengenali penyakit pada daun mangga berdasarkan ciri-ciri visualnya, seperti warna, tekstur, dan pola bercak ?

1.3 Pembatasan Masalah

1. Data citra daun mangga yang digunakan dalam penelitian diambil dari website dataset *Kaggle*.
2. Penelitian ini hanya pada beberapa jenis penyakit daun mangga yang umum, seperti bercak daun dan infeksi jamur, tanpa mencakup seluruh jenis penyakit daun mangga.

1.4 Tujuan

Penelitian ini bertujuan untuk mengembangkan model untuk deteksi penyakit daun mangga menggunakan algoritma *K-Means Clustering* yang mampu mengidentifikasi dan mengelompokkan penyakit secara efektif. Serta mampu menghasilkan proses pengelompokan citra daun mangga berdasarkan ciri visual seperti warna, tekstur, dan pola bercak dengan tingkat akurasi citra daun mangga.

1.5 Manfaat

Penelitian ini tidak hanya membantu petani dengan meningkatkan hasil pertanian mereka, tetapi juga dapat menjadi referensi untuk penelitian lebih lanjut tentang pertanian presisi dan pengolahan citra. Dengan menggunakan teknologi ini, diharapkan keberlanjutan dan produktivitas pertanian akan meningkat.

1.6 Sistematika Penulisan

Penyusunan laporan tugas akhir mengikuti tata cara penulisan sebagai berikut:

BAB 1 : PENDAHULUAN

Bab ini mencakup penjelasan tentang latar belakang, perumusan masalah, pembatasan masalah, tujuan penelitian, serta manfaat penelitian, dan juga sistematika penulisan.

BAB 2 : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini membahas penelitian-penelitian sebelumnya yang relevan serta dasar teori yang mendasari penelitian.

BAB 3 : METODE PENELITIAN

Bab ini menggambarkan tahapan penelitian yang dimulai dari pengumpulan data, pembuatan *flowchart* sistem hingga perancangan implementasi sistem.

BAB 4 : HASIL DAN ANALISIS PENELITIAN

Bab ini menyajikan hasil penelitian, yaitu segmentasi citra dengan menggunakan algoritma *k-means*.

BAB 5 : KESIMPULAN DAN SARAN

Bab terakhir menyajikan kesimpulan dan saran dari proses awal sampai akhir penelitian.

BAB II
TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Pada bagian ini, akan dibahas mengenai literatur dan penelitian terdahulu yang relevan dengan penerapan penyakit daun berbagai jenis tanaman serta penyakit medis. Penelusuran literatur sangat penting untuk memberikan landasan teori yang kuat serta menghindari duplikasi penelitian.

Beberapa penelitian salah satunya (Manalu dkk., 2023) untuk mengidentifikasi jenis penyakit pada bawang merah dengan menggunakan sampel citra dari tanaman yang terkena penyakit. Proses dimulai dengan mengubah citra tersebut menjadi format *grayscale*. Setelah itu, citra diproses menggunakan Gabor Filter untuk mengekstrak nilai magnitude. Hasil ekstraksi ini kemudian dikelompokkan ke dalam berbagai kategori penyakit bawang merah dengan memanfaatkan algoritma *K-Means Clustering*.

Berdasarkan penelitian sebelumnya yang menggunakan metode *K-Means* untuk melakukan segmentasi area yang terinfeksi penyakit pada gambar daun. Setelah menyelesaikan proses segmentasi, langkah selanjutnya adalah mengenali penyakit yang menyerang daun jeruk. Terdapat tiga jenis penyakit yang diidentifikasi, yaitu CVPD *(Citrus Vein Phloem Degeneration), Cendawan Jelaga, dan Downy Mildew*. Proses pengenalan ini dilakukan dengan memanfaatkan algoritma *K-Nearest Neighbor* (KNN). Segmentasi pada area yang terkena penyakit dilakukan untuk meningkatkan akurasi dan kualitas dalam proses klasifikasi data uji terhadap data latih menggunakan algoritma K-NN. (Febrinanto dkk., 2018).

Berdasarkan penelitian (Puerwandono & Maulana, 2023) menggunakan algoritma *Support Vector Machines* (SVM) untuk penelitian daun sirih hijau dan merah adalah jenis yang diklasifikasikan. Pertama, gambar diresize, lalu diubah menjadi *grayscale*. Segmentasi dilakukan dengan algoritma *K-Means Clustering*. Selanjutnya, parameter *eccentricity* dan metrik digunakan untuk mengekstraksi ciri. Akurasi pelatihan sebesar 91% dan akurasi pengujian sebesar 80%.

Penelitian yang menggunakan *K-Means Clustering* untuk menganalisis pola penyakit menular memberikan banyak wawasan menarik. Misalnya, ada penelitian yang menemukan bahwa kelompok (*cluster*) terbesar dalam data penyakit menular adalah diare. Hal ini menunjukkan bahwa diare mungkin menjadi masalah kesehatan yang paling dominan, yang bisa jadi disebabkan oleh faktor-faktor seperti kurangnya akses air bersih, kebersihan lingkungan, atau kondisi sanitasi yang buruk. Sementara itu, penelitian lain oleh Silitonga dan Morina menganalisis pola penyebaran penyakit berdasarkan usia pasien. Hasilnya, kelompok usia yang paling banyak terkena penyakit adalah pasien usia lanjut, diikuti oleh pasien usia muda (Mashur & Salim, 2022).

Peneliti (Premana dkk., 2020) memanfaatkan teknik segmentasi gambar digital untuk menganalisis tekstur dan warna. Dalam prosesnya, Filter Gabor digunakan untuk mengekstraksi fitur-fitur tersebut, sementara algoritma *K-Means Clustering* digunakan untuk melakukan segmentasi gambar. Tingkat keberhasilan menjadi indikator utama dalam mengukur kualitas hasil segmentasi. Untuk memastikan hasil segmentasi yang berkualitas, satu jenis gambar dari setiap sampel dataset diambil secara acak untuk diuji. Hal ini bertujuan menghasilkan gambar dengan kualitas tinggi yang dapat mendukung tahap analisis gambar berikutnya. Analisis gambar membutuhkan input yang berasal dari segmentasi dengan hasil optimal agar dapat memberikan informasi yang akurat. Penelitian ini menggunakan 50 gambar digital sebagai sampel data, yang diambil menggunakan kamera DSLR Canon EOS 60D melalui proses fotografi manual. Untuk memaksimalkan pengenalan pola, penelitian ini memanfaatkan perangkat lunak MATLAB R2017b dengan implementasi algoritma *K-means clustering* dan Filter Gabor.

2.2 Dasar Teori

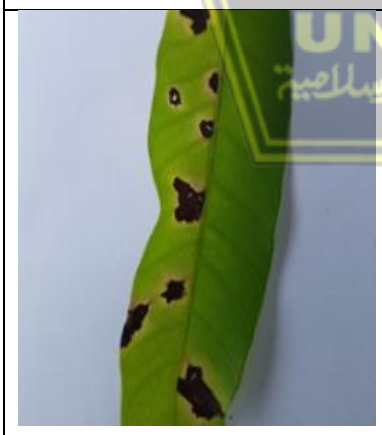
2.2.1 Penyakit Daun Mangga

Mangga atau *Mangifera Indica L.,* tanaman dari famili *Anacardiacea*, adalah salah satu jenis tanaman buah yang berasal dari India. Mangga memiliki nilai gizi tinggi yang dapat membantu meningkatkan daya tahan tubuh terhadap sariawan dan kerusakan mata. Daun adalah organ tumbuhan yang sangat penting karena di sana fotosintesis dan respirasi terjadi, serta banyak nutrisi disimpan di daun. Akibatnya, daun tanaman dapat berdampak negatif pada buahnya jika rusak (Hidayat, 2022).

Namun, tanaman mangga rentan terhadap berbagai penyakit yang dapat menyerang daun dan merusak kualitas buahnya. Penyebab penyakit tanaman mangga umumnya disebabkan oleh Parasit dan Non-Parasit. Salah satu penyebab penyakit parasit adalah Penyakit jamur (*jamur gloesoporium*) atau bakteri yang biasanya menyerang bagian akar, batang, kulit batang, ranting, buah, dan daun. Sementara itu, penyebab penyakit non-parasit meliputi faktor air, suhu, cahaya, dan nutrisi. Beberapa penyakit yang umum ditemukan pada daun mangga antara lain bercak daun, busuk daun, dan infeksi jamur (Solikin, 2020)

Penyakit yang umum muncul pada daun mangga adalah *klorosis*. *Klorosis* merupakan kondisi di mana jaringan tumbuhan, khususnya bagian daun, mengalami perubahan warna dari hijau ke kuning akibat kekurangan *klorofil*. Perubahan warna yang terjadi disebabkan oleh *klorofil* (zat hijau daun) yang mengalami kerusakan atau cacat (Rosadi dkk., 2023). Penyebabnya adalah penyakit non-parasit dan kondisi fisiologis, seperti kekurangan atau kelebihan unsur hara, air, sinar matahari, dan suhu. Bercak daun pada mangga biasanya disebabkan oleh jamur seperti *Corynespora cassiicola* dan *Cercospora spp.,* yang menyebabkan munculnya bercak-bercak hitam atau coklat pada daun. Busuk daun sering disebabkan oleh jamur *Fusarium* dan *Phytophthora*, yang mengakibatkan daun menjadi coklat dan membusuk. Infeksi jamur ini dapat menyebar dengan cepat, dan jika tidak ditangani dengan baik, bisa berdampak besar terhadap hasil panen.

Tabel 2. 1 Jenis Penyakit Daun Mangga

Gambar Daun	Jenis Penyakit	Disebabkan Oleh
	Tidak memiliki penyakit (sehat)	
	*Bacterial Canker*	Bakteri yang menyebabkan area mati pada daun berupa lesi yang melingkar atau lonjong, dan dapat berubah warna.
	*Antracnose*	Penyakit jamur yang menyebabkan lesi gelap pada daun tanaman

	*Gall Midge*	Penyakit yang disebabkan oleh lalat empedu yang bertelur di daun.
	*Sooty Mould*	disebabkan oleh jamur jelaga yang menyebabkan pertumbuhan jamur berwarna hitam atau coklat tua pada permukaan daun
	*Powdery Mildew*	Penyakit yang disebabkan oleh jamur *Oidium mangiferae*

2.2.2 *K-Means Clustering*

Metode *K-Means* diperkenalkan oleh *MacQueen JB* pada tahun 1976. Tujuan dari metode ini adalah untuk mengelompokkan objek berdasarkan atributnya ke dalam k partisi. *K-Means* sangat populer karena kemudahan penggunaannya dan kemampuannya dalam mengklaster data besar serta mengidentifikasi dengan cepat (Mashur & Salim, 2022).

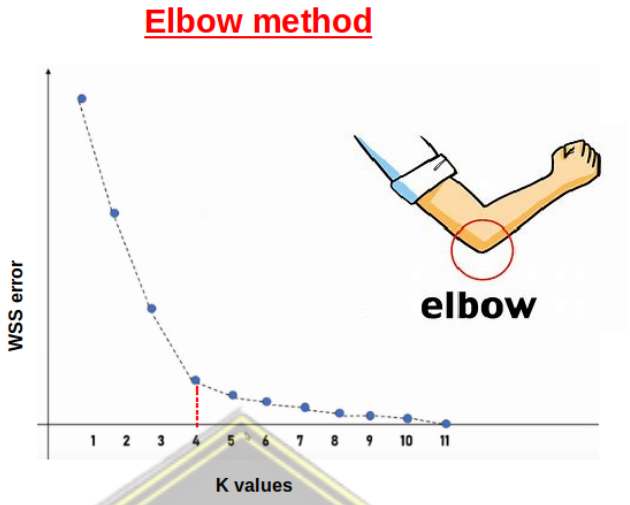
Salah satu langkah penting dalam penerapan metode *K-Means* adalah penentuan centroid, jumlah cluster, dan jarak antar centroid. Dengan membentuk beberapa cluster menggunakan *K-Means*, kita juga dapat mengukur jarak antara pusat *cluster (centroid)* dalam data yang sedang dianalisis. Hasil ini menjadi dasar untuk mengklasifikasikan data baru yang muncul, sehingga kita dapat mengetahui kelompoknya.

Algoritma *K-Means* merupakan salah satu metode pengelompokan tidak bersifat hierarki yang bertujuan untuk membagi data menjadi satu atau lebih kluster. Metode ini termasuk dalam kategori pembelajaran tanpa pengawasan *(unsupervised learning)*, di mana algoritma ini berfungsi tanpa memerlukan label atau kategori data yang sudah ditentukan sebelumnya.

Segmentasi gambar adalah proses mempartisi gambar menjadi berbagai kelompok yang memiliki fitur atau kesamaan tertentu. Dalam hal ini, K-Means digunakan untuk mengelompokkan gambar sedemikian rupa sehingga setiap kluster setidaknya terdiri dari gambar-gambar yang memiliki kesamaan area utama yang relevan.

Clustering adalah suatu tahapan untuk mengelompokkan data ke dalam beberapa kategori, sehingga data yang berada dalam satu kluster memiliki tingkat kesamaan yang tinggi, sementara perbedaan antara kluster satu dengan yang lainnya diminimalkan. Metode *clustering* telah diterapkan di berbagai bidang untuk melakukan segmentasi data (Suroyo, 2019).

Cara menentukan K pada *K-Means* yaitu menggunakan metode yang umum digunakan untuk menentukan nilai k yaitu Metode *Elbow*. Metode ini adalah cara yang paling populer untuk memilih jumlah *cluster* yang optimal. Ide dasarnya adalah untuk *memplot Inertia* disebut juga *Within-Cluster Sum Of Squares (WCSS)* terhadap jumlah cluster K dan melihat di mana terdapat "siku" *(elbow)* pada grafik.



Gambar 2. 1 Grafik *WSS*

Misalkan saya mencoba beberapa nilai K dari 1 hingga 10 dan menghitung inertia-nya. Jika grafik menunjukkan penurunan inertia yang tajam dari K=1 hingga K=4, tetapi mulai datar setelah K=4, maka K=4 bisa dianggap sebagai jumlah cluster yang optimal.

Metode perhitungan jarak dalam penerapan *K-Means* menggunakan *Euclidean Distance*. Metode ini digunakan untuk mengukur jarak antara dua titik dalam ruang *Euclidean*. Formula *Euclidean Distance* antara dua titik $(x1, y1)$ dan $(x2, y2)$ dalam ruang dua dimensi adalah:

$$d = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$

Langkah-langkah Algoritma *K-Means* adalah sebagai berikut:

1. Menentukan jumlah cluster (k) yang diinginkan.
2. Menentukan titik pusat atau centroid untuk masing-masing cluster diukur menggunakan *Euclidean distance.*
3. Hitung *centroid* baru berdasarkan rata-rata posisi data dalam *cluster*.
4. Proses ini diulang hingga centroid tidak berubah lagi *(konvergen).*

2.2.3 *Silhoutte Score*

Silhoutte Score adalah metrik yang digunakan untuk mengevaluasi kualitas klaster dalam clustering. Metrik ini memberikan informasi tentang seberapa baik data dalam suatu klaster dikelompokkan dibandingkan dengan klaster lainnya.

Skor ini berkisar dari -1 hingga 1:

a. Nilai mendekati +1: Titik data dikelompokkan dengan benar.
b. Nilai mendekati 0: Titik data berada di perbatasan antara dua klaster.
c. Nilai mendekati -1: Titik data dikelompokkan ke klaster yang salah.

Untuk setiap titik data i, *Silhouette Score* dapat dihitung sebagai :

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

$a(i)$: Rata-rata jarak i ke semua titil lain dalam clusternya *(intra-cluster distance)*

$b(i)$: Rata-rata jarak i ke semua titil lain dalam cluster terdekat *(inter-cluster distance)*

Skor keseluruhan adalah rata-rata semua s(i) dimana n adalah jumlah data :

$$\textit{Sillhoutte Score} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{s(i)}$$

Langkah-langkah Perhitungan *Silhoutte Score* :

1. Hitung jarak antar semua titik, untuk menghitung jarak antar 2 titik, digunakan rumus *Eucledian Distance* :

   $d(x1, x2) = \sqrt{(x1 - x2)^2 + (y1 - y2)^2}9$ (1)

2. Hitung $a(i)$ adalah rata-rata jarak titik ke i ke semua titi lain dalam clusternya. Misal perhitungan untuk pelanggan A (Cluster 1)
3. Hitung $b(i)$ adalah rata-rata jarak titik i ke semua titik dalam clauster terdekat. Misal Cluster 1 dengan cluster terdekat adalah cluster 2
4. Hitung *Silhoutte Score* $s(i)$

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (2)$$

5. Hitung *Silhouette Score* Keseluruhan

   Rata-rata $s(i)$ dari semua titik :

   Sillhoutte Score $= \frac{1}{n} \sum_{i=1}^{n} \mathrm{s(i)}$ (3)

2.2.4 *Machine Learning*

Pembelajaran Mesin *(Machine Learning*) adalah sebuah teknologi yang dirancang untuk memungkinkan mesin belajar secara mandiri tanpa perlu bimbingan dari penggunanya. Konsep pembelajaran mesin ini dibangun di atas berbagai disiplin ilmu, seperti statistik, matematika, dan penambangan data (data mining), sehingga mesin dapat belajar dengan menganalisis data yang ada tanpa harus diprogram ulang atau diberikan perintah secara langsung. Dalam hal ini, machine learning memiliki kemampuan untuk mengolah data berdasarkan instruksi yang dihasilkan sendiri.

Peran *machine learning* sangat krusial dalam membantu manusia di berbagai bidang. Metode kerja *machine learning* sebenarnya bervariasi, tergantung pada teknik atau pendekatan pembelajaran yang digunakan. Namun, secara umum, prinsip dasar cara kerja pembelajaran mesin tetap konstan, yang mencakup pengumpulan data, eksplorasi data, pemilihan model atau teknik yang sesuai, pelatihan model yang telah dipilih, serta evaluasi hasil yang diperoleh dari proses *machine learning* tersebut.

Secara luas *Machine Learning* memiliki dua teknik dasar belajar, yaitu *:*

1. *Supervised Learning*

   Supervised learning merupakan teknik pada *machine learning* untuk memberikan label tertentu.

2. *Unsupervised Learning*

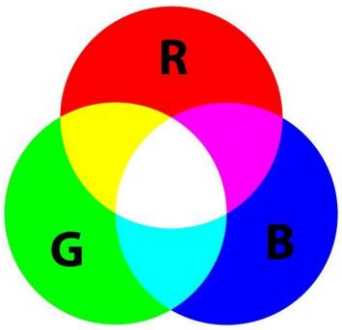
   Unsupervised learning merupakan teknik pada *machine learning* untuk membantu menemukan struktur atau pola tersembunyi pada data yang tidak memiliki label.

2.2.5 *Image Processing*

Pengolahan citra, atau yang dikenal dengan istilah *Image Processing*, adalah suatu sistem di mana proses dilakukan dengan gambar sebagai masukan (*input*), dan hasilnya juga berupa gambar sebagai keluaran (*output*) (Mulyawan dkk., 2019). Awalnya, pengolahan gambar ini berfokus pada peningkatan kualitas visual gambar. Namun, seiring dengan kemajuan dalam dunia komputasi terutama yang ditandai dengan peningkatan kapasitas dan kecepatan proses computer serta lahirnya ilmu komputer yang memungkinkan kita untuk menganalisis dan mendapatkan informasi dari gambar, pengolahan citra kini tidak dapat dipisahkan dari bidang visi komputer atau *computer vision*.

Model Citra :

1. *Grayscale,* disebut sebagai *grayscale cycling,* digunakan untuk meningkatkan kualitas visual gambar dengan cara mengatur kecerahan dan kontrasnya. Melalui metode ini, gambar dapat dimodifikasi menjadi lebih menarik dan lebih nyaman untuk dilihat (Pramudiya *dkk.*, 2024).
2. *RGB (Red, Green, Blue)* terdiri dari tiga bidang citra yang terpisah, masing-masing mewakili warna dasar: merah, hijau, dan biru pada setiap piksel (Agyztia Premana dkk., 2020).



Gambar 2. 2 RGB

Untuk mengubah gambar berwarna penuh (RGB) menjadi citra *grayscale* (abu-abu), rumus yang digunakan sebagai berikut :

$$(R + G + B)/3$$

Dimana :

R : Unsur warna merah

G : Unsur warna hijau

B : Unsur warna biru

Nilai yang dihasilkan dari persamaan tersebut akan dimasukkan ke dalam masing-masing unsur warna dasar citra *grayscale*.

2.2.6 Deteksi Tepi *Canny*

Salah satu operator yang efektif dalam mendeteksi citra tepi daun adalah operator *Canny*. Operator ini dikenal sebagai salah satu metode deteksi tepi yang paling optimal (Budianita dkk., 2019). *Canny Edge Detector* adalah operator deteksi tepi yang memanfaatkan algoritma multi-tahap untuk mendeteksi berbagai tepi dalam gambar. Metode ini dikembangkan oleh *John F. Canny* pada tahun 1986 terkenal sebagai operator deteksi tepi yang optimal. *Canny* juga menghasilkan teori komputasi deteksi tepi yang menjelaskan mengapa teknik ini bekerja.

Dalam segmentasi gambar, algoritma deteksi tepi *Canny* melalui beberapa langkah yang berurutan. Pertama, algoritma ini mengurangi kebisingan pada gambar dengan menerapkan filter Gaussian. Selanjutnya, menghitung gradien untuk mengidentifikasi perubahan intensitas. Setelah itu, dilakukan penekanan non-maksimum untuk memastikan bahwa hanya lokal maksimum yang diperhitungkan sebagai tepi. Terakhir, tepi-tepi tersebut dipantau menggunakan metode histeresis untuk memastikan keakuratan dan konsistensi hasil deteksi (Sinra dkk., 2023).

Berikut penjelasan langkah-langkahnya

1. *Gaussian Blur*

   Salah satu metode untuk mengurangi noise pada gambar adalah dengan menggunakan *Gaussian blur* untuk memperhalusnya. Agar dapat melakukannya, teknik konvolusi citra diterapkan menggunakan *Kernel*

Gaussian (3x3, 5x5, 7x7 dan seterusnya…). Pada dasarnya, kernel yang paling kecil, yang kurang terlihat adalah blur.

2. Gradien Intensitas Gambar

   Tepi pada gambar dapat mengarah ke berbagai arah, sehingga algoritma *Canny* menerapkan empat filter untuk mendeteksi tepi horizontal, vertikal, dan diagonal pada gambar yang buram. Operator deteksi tepi (seperti *Roberts, Prewitt,* atau *Sobel*) menghasilkan nilai untuk turunan pertama pada arah horizontal (Gx) dan arah vertikal (Gy). Dari sini, gradien dan arah tepi dapat diidentifikasi : $G \sqrt{G_x^2 + Gy^2}$

3. Penekanan *non-maximum*

   Penekanan batas besaran gradien adalah teknik penipisan tepi. Penekanan batas diterapkan untuk menemukan lokasi dengan perubahan nilai intensitas paling tajam.

   Algoritma untuk setiap piksel dalam gambar gradien yaitu :

   a. Bandingkan kekuatan tepi piksel saat ini dengan kekuatan tepi piksel dalam arah gradien positif dan negatif.
   b. Jika kekuatan tepi piksel saat ini adalah yang terbesar dibandingkan dengan piksel lain dalam topeng dengan arah yang sama (misalnya, piksel yang menunjuk ke arah y akan dibandingkan dengan piksel di atas dan di bawahnya pada sumbu vertikal), nilainya akan dipertahankan. Jika tidak, nilainya akan ditekan.

4. Ambang Batas

   Setelah penerapan penekanan *non-maksimum*, piksel tepi yang tersisa memberikan representasi tepi yang nyata lebih akurat dalam gambar. Jika nilai gradien piksel tepi melebihi nilai ambang batas tinggi, maka itu ditandai sebagai piksel tepi yang kuat. Jika nilai gradien piksel tepi kurang dari nilai ambang batas tinggi dan lebih besar dari nilai ambang batas rendah, maka itu ditandai sebagai piksel tepi yang lemah. Jika nilai gradien piksel tepi kurang dari nilai ambang batas rendah, maka itu akan ditekan. Kedua nilai ambang batas ditentukan secara empiris dan definisinya akan tergantung pada konten gambar yang diberikan.

5. Pelacakan Tepi dengan Histeresis.

   Piksel tepi yang kuat harus terlibat dalam gambar tepi akhir, piksel tersebut dianggap berasal dari tepi sebenarnya dalam gambar. Algoritma ini menggunakan piksel tepi yang lemah dari tepi sebenarnya yang terhubung ke piksel tepi yang kuat.

2.2.7 *Python*

Python merupakan bahasa pemrograman tingkat tinggi yang populer karena sintaksnya yang mudah dan simpel, sehingga memudahkan pengguna untuk menulis serta memahami kode. Diciptakan oleh *Guido van Rossum* dan pertama kali dirilis pada tahun 1991, *Python* mendukung berbagai paradigma dalam pemrograman, seperti pemrograman berorientasi objek, pemrograman fungsional, serta pemrograman terstruktur. *Python* digunakan secara luas di berbagai sektor seperti pengembangan web, sains data, pembelajaran mesin, otomatisasi skrip, dan analisis data.

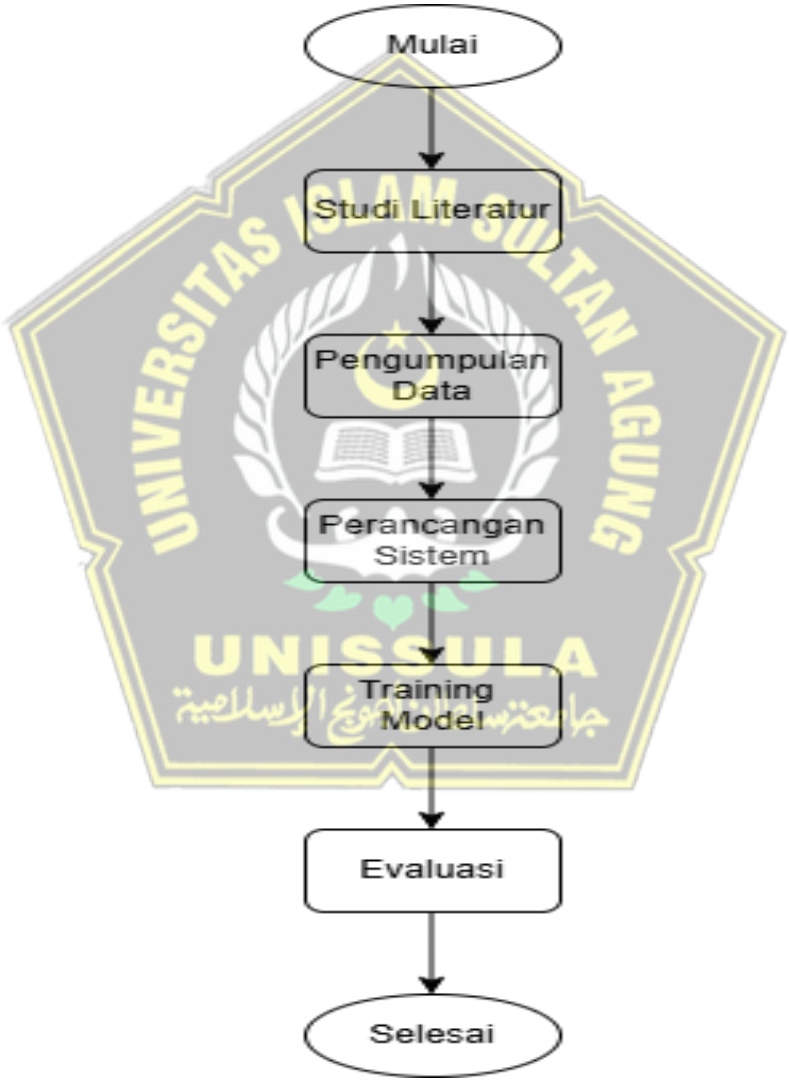
Gambar 2. 3 Logo *Python*

Salah satu keunggulan *Python* adalah ketersediaan pustaka yang luas dan kuat, yang memungkinkan pengguna untuk menangani berbagai macam tugas secara efisien. Beberapa pustaka terkenal meliputi *NumPy* dan *Pandas* untuk manipulasi data, *Matplotlib* dan *Seaborn* untuk visualisasi data, serta *TensorFlow* dan Keras untuk pembelajaran mesin dan kecerdasan buatan. *Python* juga memiliki ekosistem yang mendukung pengembangan aplikasi yang bersifat *multi-platform*, sehingga dapat digunakan di berbagai sistem operasi, seperti *Windows, macOS, dan Linux*. *Python* sangat populer di kalangan pemula maupun profesional karena kemudahan dalam belajar dan penggunaannya yang fleksibel.

BAB III
METODE PENELITIAN

3.1 Metode Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang sistematis untuk mencapai tujuan yang telah ditetapkan terhadap algoritma *k-means clustering* untuk segmentasi penyakit daun mangga.



Gambar 3. 1 *Flowchart* Alur Sistem

3.2 Studi Literatur

Studi literatur dilakukan dengan mengumpulkan dan menelaah artikel serta jurnal ilmiah dari penelitian-penelitian terdahulu yang berhubungan dengan penulisan tugas akhir. Dari daftar pustaka yang dikaji, penelitian ini mengacu pada jurnal akademik dari berbagai institusi dan penerbit seperti *Elsevier*, IEEE Xplore, dan *Google Scholar,* serta jurnal seperti Jurnal Buana Informatika, JATISI, dan Jurnal Komputer Terapan. Selain itu, dataset diperoleh dari website seperti *Kaggle* untuk digunakan dalam penulisan tugas akhir sehingga membantu peneliti dalam menyelesaikan penelitian ini.

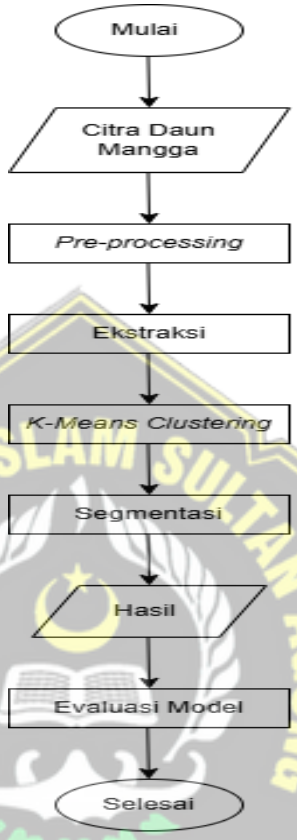
3.3 Pengumpulan Data

Kualitas dan keberagaman data akan sangat memengaruhi hasil segmentasi algoritma *K-Means*, sehingga pengumpulan data merupakan langkah penting dalam penelitian ini. Peneliti menggunakan dataset sekitar 2000 foto daun mangga yang sehat dan terinfeksi penyakit, yang diambil dari salah satu sumber data terbuka terbesar di dunia yaitu website *Kaggle*. Data ini mencakup berbagai kondisi kesehatan, seperti *Anthracnose, Bacterial Canker, Powder Mildey* dan penyakit lainnya.

3.4 Perancangan Implementasi Sistem

Perancangan dan implementasi sistem bertujuan untuk mengelompokkan citra daun mangga berdasarkan tingkat keparahan penyakitnya dengan memanfaatkan algoritma *k-means clustering*. Pada tahap ini juga merancang modul untuk *preprocessing, clustering*, dan visualisasi hasil. Pemilihan jumlah klaster (nilai k) dilakukan melalui pendekatan eksperimen agar dapat memperoleh hasil pengelompokan yang optimal. Hasil dari proses perancangan ini menjadi landasan bagi implementasi sistem berbasis Python. Sistem dirancang agar dapat beroperasi

secara otomatis dan menyediakan visualisasi hasil segmentasi, sehingga memudahkan pengguna dalam menginterpretasikan kondisi daun mangga.



Gambar 3. 2 *Flowchart* Perancangan Sistem

a. Proses *preprocessing* yaitu setiap gambar daun ukurannya menjadi ukuran yang standar, misalnya 256x256 *piksel* untuk melakukan pengolahan awal pada citra yang dikumpulkan untuk meningkatkan kualitas gambar. Selanjutnya mengubah gambar ke format *grayscale* untuk memudahkan analisis langkah ini meliputi penyesuaian kontras, pengurangan noise, dan resizing gambar.
b. Tahap Ekstraksi Peneliti fokus pada pencarian ciri-ciri yang menunjukkan ada atau tidaknya penyakit pada daun mangga, seperti bentuk dan warna daun. Ciri-ciri ini akan digunakan untuk mengelompokkan daun-daun yang sehat dan yang terinfeksi.
c. *K-Means Clustering* yaitu inti dari penelitian menggunakan algoritma *K-Means* untuk mengelompokkan daun-daun berdasarkan ciri-ciri yang sudah

ditemukan. *K-Means* akan mencoba memisahkan daun berdasarkan data yang ada. Data yang telah diproses dikelompokkan menggunakan algoritma *K-Means Clustering* berdasarkan kemiripan ciri visualnya.

d. Perhitungan segmentasi gambar hanya memanfaatkan dua variabel dari ruang warna L*a*b*, yaitu variabel a* dan b*. Segmentasi daun, yang memisahkan bagian daun dari tutup, serta segmentasi penyakit, yang memisahkan bagian penyakit dari daun, merupakan dua jenis segmentasi yang digunakan untuk memisahkan kemurnian warna.

e. Hasil Klasifikasi, sistem memberikan hasil berupa diperoleh dengan menganalisis citra yang telah diproses.

f. Evaluasi model bertujuan untuk mengukur kinerja model menggunakan beberapa metrik utama, seperti akurasi, presisi, *recall*, dan *f1-score.*

1. Akurasi yaitu mengukur seberapa banyak piksel yang diklasifikasikan dengan benar dibandingkan seluruh piksel.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2. Presisi yaitu mengukur sejauh mana semua piksel yang diprediksi sakit, berapa banyak yang benar-benar sakit.

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

3. *Recall* yaitu mengukur sejauh mana piksel yang benar-benar sakit, berapa banyak yang berhasil dideteksi.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

4. *F1-score* yaitu rata-rata harmonik dari presisi dan recall yang digunakan untuk menyeimbangkan keduanya.

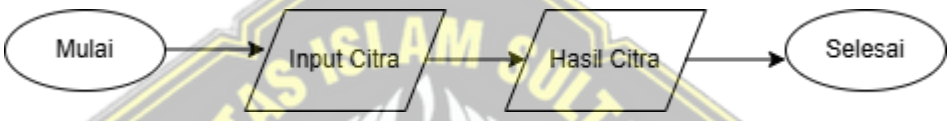
$$FI\text{-}score = 2 \times \frac{presisi \times recall}{presisi + recall} \quad (4)$$

Keterangan:

- TP (*True Positive*) = Prediksi piksel sakit.
- TN (*True Negative)* = Prediksi piksel sehat.
- FP (*False Positive)* = Piksel sehat tapi diprediksi sakit.
- FN (*False Negative)* = Piksel sakit tapi diprediksi sehat.

3.5 Pengujian Sistem

Tahap ini merupakan implementasi sistem yang sudah dirancang sebelumnya. Alur kerja sistemnya dapat ditemukan pada Gambar 3.3



Gambar 3. 3 *Flowchart* Pengujian Sistem

Flowchart ini menggambarkan langkah-langkah dalam proses pengujian sistem pengolahan citra. Proses dimulai dari tahap Mulai, yang berfungsi sebagai titik awal pelaksanaan pengujian. Selanjutnya, sistem melakukan Input Citra, dimana gambar yang akan diuji dimasukkan ke dalam sistem. Setelah citra berhasil dimasukkan, sistem melanjutkan dengan memprosesnya untuk menghasilkan hasil citra yaitu hasil dari pengolahan atau analisis yang dilakukan terhadap citra tersebut. Setelah hasil citra diperoleh, proses berlanjut ke tahap Selesai, yang menandakan bahwa seluruh rangkaian pengujian sistem telah selesai dilaksanakan. *Flowchart* ini memiliki alur yang linier, menyajikan proses kerja yang sederhana namun jelas dalam konteks pengujian sistem berbasis citra.

BAB IV
HASIL DAN ANALISIS PENELITIAN

4.1 Data Citra

Data citra yang digunakan dalam penelitian ini diperoleh dari platform *Kaggle*, yang menyediakan berbagai dataset yang berkaitan dengan penyakit daun mangga. Dataset ini termasuk gambar daun sehat dan terinfeksi yang memungkinkan penggunaan algoritma *K-Means* untuk segmentasi. Penelitian ini bertujuan untuk meningkatkan akurasi identifikasi penyakit pada daun mangga dengan menggunakan gambar yang beragam.

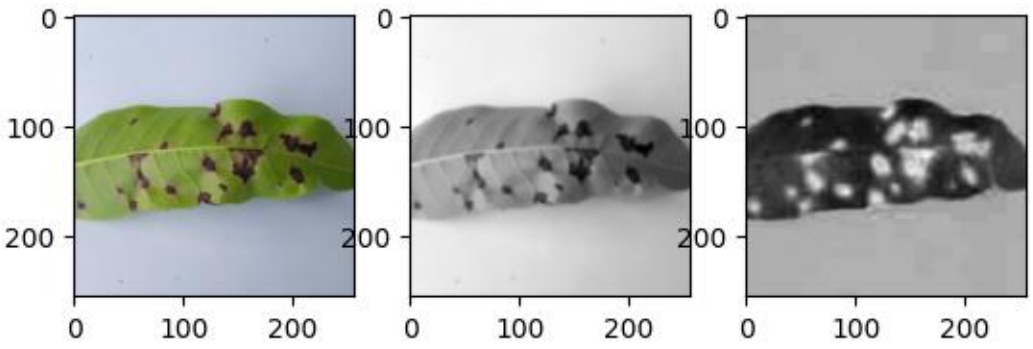
Lebih dari 2000 foto daun mangga, yang diambil dari data *Kaggle*, dibagi menjadi dua kelas utama: daun sehat dan daun yang terinfeksi penyakit seperti *Anthracnose, Bacterial Canker, Powder Mildey* dan penyakit lainnya. Gambar-gambar dalam dataset memiliki resolusi yang berbeda, sebagian besar memiliki kualitas yang sangat baik, yang penting untuk analisis yang akurat. Variasi dalam gambar ini mencakup berbagai kondisi pencahayaan dan sudut pengambilan gambar, yang membuat proses segmentasi menjadi sulit.

Untuk meningkatkan jumlah data dan variasi, teknik augmentasi citra diterapkan. Tujuan dari augmentasi ini adalah untuk meningkatkan kemampuan model untuk mengidentifikasi pola dan karakteristik penyakit pada daun mangga. Algoritma *K-Means* digunakan dalam segmentasi citra untuk membedakan area yang terinfeksi dari area yang sehat. Diharapkan algoritma ini dapat menghasilkan hasil segmentasi yang akurat dan efektif dalam mendeteksi penyakit pada daun mangga dengan menggunakan data yang beragam ini.

4.2 *Preprocessing* Citra

Preprocessing gambar, langkah penting dalam pengolahan gambar yang bertujuan untuk meningkatkan kualitas gambar sebelum diproses dalam penelitian ini untuk meningkatkan kontras, dan menstandarisasi ukuran gambar. Mengubah gambar berwarna menjadi gambar *grayscale* adalah bagian penting dari proses *preprocessing*. Proses ini terdiri dari beberapa tahapan:

1. Konversi ke *grayscale* : gambar awal daun mangga menjadi *grayscale*. Mengingat algoritma *K-Means* lebih efektif dalam mengelompokkan gambar dengan satu saluran warna, proses konversi ini dilakukan untuk menyederhanakan data yang akan diproses. Nilai intensitas untuk setiap piksel dalam gambar *grayscale* berkisar antara 0 dan 255, yang menunjukkan warna hitam. Mengurangi dimensi warna membuat proses segmentasi lebih cepat dan efisien dan memudahkan analisis fitur.
2. Pengubahan Ukuran Citra: Setelah citra diubah menjadi *grayscale*, setiap gambar diubah menjadi ukuran yang seragam, yaitu 256 x 256 piksel. Pengubahan ukuran ini dilakukan untuk menjaga konsistensi dalam analisis dan memudahkan proses pemrosesan gambar, dan algoritma *K-Means* dapat mengelompokkan piksel dengan lebih efisien dengan ukuran yang seragam.
3. Pengurangan *Noise* : filter *Gaussian blur* digunakan untuk mengurangi *noise* dalam gambar melakukannya dengan menghitung rata-rata piksel di sekitar piksel yang dimaksud dapat menghasilkan gambar yang lebih halus dengan lebih sedikit detail yang tidak diinginkan. Akurasi segmentasi algoritma *K-Means* akan meningkat sebagai hasil dari pengurangan noise ini.
4. Normalisasi Citra : Setelah proses pengurangan noise, citra dinormalisasi untuk memastikan nilai intensitas piksel berada dalam rentang yang sama. Hal ini penting untuk menghindari algoritma *K-Means* dari bias yang terjadi dalam pengelompokan piksel, dapat terjadi jika ada perbedaan besar dalam skala nilai intensitas.



Gambar 4. 1 *Preprocessing* Gambar

Gambar diatas adalah tampilan proses gambar daun yang memiliki penyakit *Anthracnose* dari normal konversi ke *grayscale* untuk analisis tekstur setelah itu pengurangan *noise gaussian blur* untuk pengurangan *noise* dalam gambar selanjutnya konversi ke ruang warna L*a*b. Dimana :

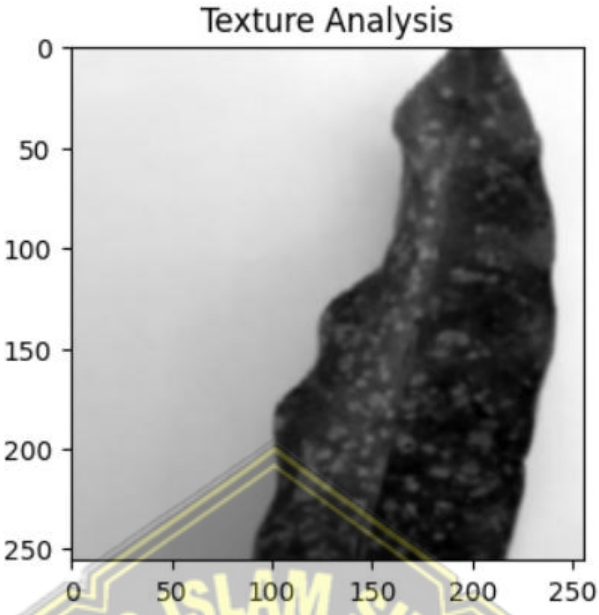
1. Gambar pertama di sisi kiri menampilkan citra asli dari sebuah daun yang menunjukkan tanda-tanda penyakit. Secara visual, daun terlihat berwarna hijau dengan sejumlah bercak gelap yang menyebar di permukaannya. Bercak-bercak tersebut biasanya menjadi indikasi adanya infeksi patogen, seperti jamur, bakteri, atau virus, yang mengakibatkan kerusakan pada jaringan daun. Identifikasi awal terhadap bercak semacam ini sangat penting dilakukan agar penyebaran penyakit dapat dicegah dan penanganan dapat diberikan tepat waktu.
2. Gambar kedua yang terletak ditengah, menunjukkan gambar yang telah dikonversi menjadi *grayscale* (abu-abu). Pada versi ini, Konversi ini dilakukan untuk menyederhanakan data warna yang rumit menjadi nilai intensitas cahaya saja, tanpa kehilangan informasi penting mengenai bentuk dan distribusi bercak penyakit. Dalam skala abu-abu, setiap piksel direpresentasikan dengan nilai intensitas antara hitam (0) hingga putih (255), sehingga perbedaan kontras antara bagian sehat dan bagian yang terinfeksi menjadi lebih mudah untuk dianalisis sebelum dilakukan segmentasi lebih lanjut.

3. Gambar ketiga, di sisi kanan, menunjukkan hasil pengolahan lebih lanjut dari citra skala abu-abu melalui teknik segmentasi atau *thresholding*. Pada tahap ini, sistem melakukan pemisahan antara area yang terindikasi terkena penyakit dan area yang masih sehat berdasarkan intensitas piksel. Area yang terinfeksi biasanya ditunjukkan dengan warna yang lebih terang atau menyala, sedangkan area sehat tampak lebih gelap. Proses ini sangat membantu dalam mengukur tingkat kerusakan daun, seperti menghitung luas area yang terinfeksi, bentuk bercak, dan distribusinya di seluruh permukaan daun. Informasi ini nantinya dapat digunakan untuk menentukan jenis penyakit dengan lebih spesifik atau untuk mengukur tingkat keparahan infeksi.

4.3 Ekstraksi Fitur

Pada tahap ekstraksi fitur, informasi penting dari citra daun diambil untuk dijadikan input dalam proses klasifikasi. Salah satu jenis fitur yang berpengaruh besar dalam membedakan jenis penyakit pada daun adalah fitur tekstur. Analisis tekstur bertujuan untuk menangkap pola-pola visual halus pada permukaan daun, seperti bintik-bintik, kerutan, atau variasi warna yang tidak merata, yang sering menjadi indikator utama adanya infeksi atau gangguan fisiologis pada tanaman. Metode yang umum digunakan dalam analisis tekstur adalah pendekatan berbasis filter konvolusi, seperti filter Gabor. Analisis tersebut dilakukan pada citra yang telah dikonversi menjadi format *grayscale*, agar pola tekstur dapat lebih mudah diidentifikasi secara numerik.

Gambar di bawah ini menampilkan hasil proses ekstraksi tekstur pada salah satu contoh citra daun. Dalam visualisasi ini, pola distribusi tekstur yang tampak pada permukaan daun menjadi lebih jelas setelah melalui pengolahan dengan pendekatan tekstur. Hal ini memungkinkan sistem untuk mendeteksi perbedaan visual yang mungkin tidak terlihat langsung oleh mata manusia.



Gambar 4. 2 Ektraksi Fitur

Dari gambar hasil analisis tekstur tersebut, terlihat bahwa kawasan permukaan daun yang mengalami perubahan atau kerusakan memiliki distribusi tekstur yang lebih kontras dibandingkan dengan area yang sehat. Fitur-fitur ini kemudian dikodekan dalam bentuk vektor numerik untuk dianalisis lebih lanjut oleh model klasifikasi. Nilai dari fitur tekstur seperti kontras, homogenitas, entropi, dan energi berperan penting dalam membedakan antar kelas penyakit. Oleh karena itu, ekstraksi tekstur menjadi salah satu tahap penting dalam rangkaian pengolahan citra, karena dapat meningkatkan kemampuan model dalam mengidentifikasi pola-pola penyakit dengan lebih akurat.

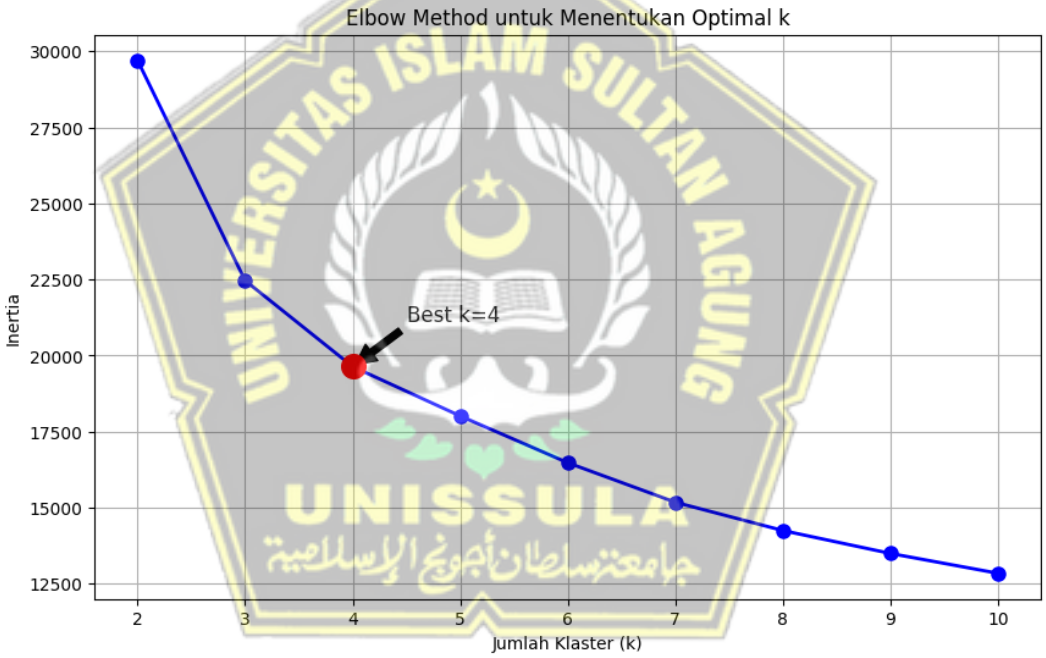
4.4 Implementasi Algoritma *K-Means Clustering*

4.4.1 *Elbow Method*

Salah satu kesulitan utama dalam menerapkan algoritma *k-means clustering* adalah menentukan jumlah kluster (k) yang paling sesuai. Jumlah kluster yang tidak tepat dapat mengakibatkan hasil pengelompokan yang tidak akurat. Jika kluster yang ditentukan terlalu sedikit, maka data yang beragam mungkin akan bergabung dalam satu kluster.

Untuk menangani masalah ini, digunakan pendekatan visual bernama Metode *Elbow*. Metode ini berfungsi untuk menemukan jumlah kluster optimal dengan cara menghitung dan memplot nilai inertia untuk berbagai nilai k.

Inertia mengacu pada jumlah kuadrat jarak antara setiap data titik dan centroid dari kluster masing-masing. Nilai inertia yang lebih rendah menunjukkan bahwa kluster tersebut lebih padat. Namun, setiap penambahan jumlah kluster akan selalu menghasilkan penurunan pada inertia, sehingga diperlukan sebuah titik kompromi yakni titik dimana penambahan kluster lebih lanjut tidak lagi memberikan pengurangan inertia yang signifikan. Titik ini dikenal sebagai "elbow".



Gambar 4. 3 Grafik *Elbow Method*

Pada grafik Metode *Elbow* di atas, tampak jelas bahwa nilai inertia menurun drastis dari k = 2 hingga k = 4. Setelah mencapai k = 4, penurunan inertia mulai melambat, yang menunjukkan bahwa menambah jumlah klaster lebih dari k = 4 tidak memberikan penurunan inertia yang berarti. Oleh karena itu, titik *elbow* ditetapkan pada k = 4, yang ditunjukkan dengan titik merah di grafik.

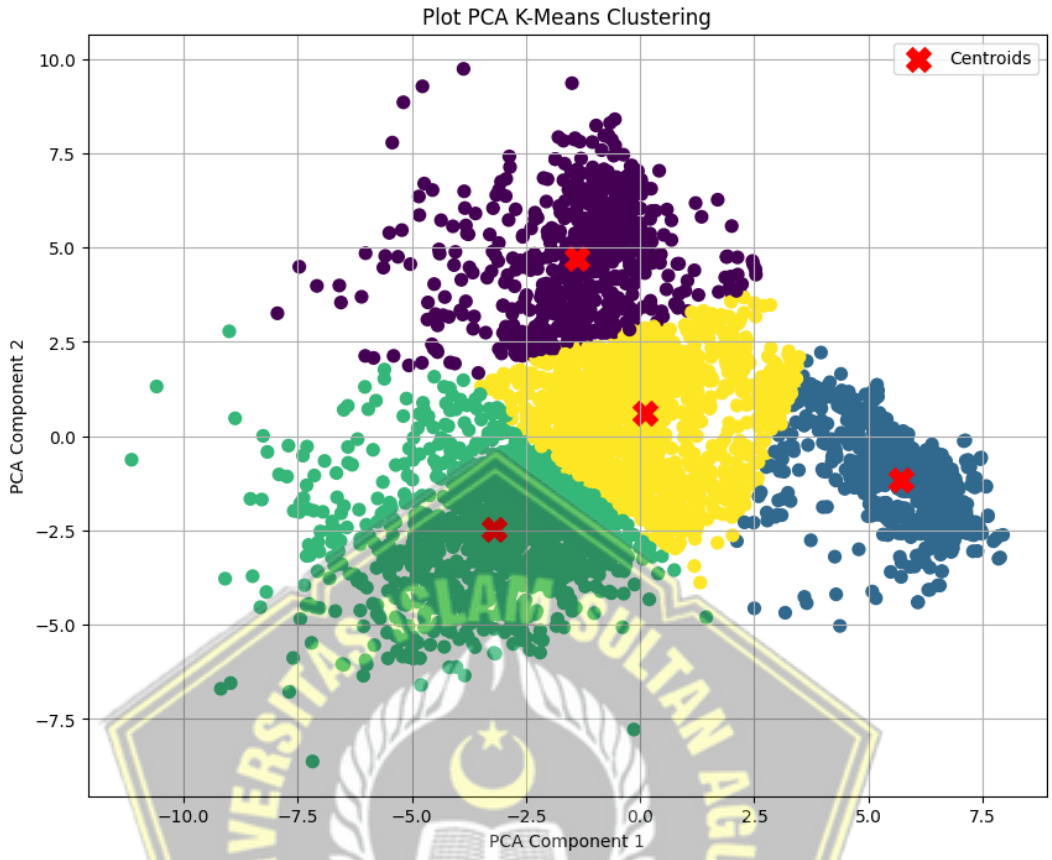
Pemilihan k = 4 sebagai jumlah klaster terbaik dalam penelitian ini mencerminkan titik efisiensi tertinggi dalam proses pengelompokan data citra daun mangga. Dengan empat klaster, data dapat dikelompokkan secara representatif tanpa meningkatkan kompleksitas yang tidak perlu. Klaster ini kemudian dimanfaatkan sebagai dasar untuk analisis lebih lanjut untuk mengidentifikasi segmen-segmen penyakit daun mangga berdasarkan karakteristik tertentu seperti warna, tekstur, atau bentuk, yang telah diperoleh melalui tahap ekstraksi fitur. Langkah ini merupakan bagian penting dalam membangun sistem klasifikasi atau deteksi penyakit berbasis kecerdasan buatan dengan akurasi dan efisiensi tinggi.

4.4.2 Visualisasi PCA dari *cluster*

Pada tahap ini, dilakukan pengelompokan menggunakan metode *K-Means Clustering* untuk mengatur data gambar daun berdasarkan kesamaan fitur warna. Fitur warna ini menunjukkan sebaran intensitas warna dalam sebuah gambar dan sangat berguna untuk mengenali pola atau karakteristik tertentu, seperti tanda-tanda penyakit pada daun.

Untuk menangani masalah ini, digunakan teknik *Principal Component Analysis* (PCA) sebagai teknik untuk mengurangi dimensi. PCA berfungsi dengan memproyeksikan data berdimensi tinggi ke dalam ruang berdimensi lebih rendah (dua dimensi). Dengan cara ini, hasil pengelompokan dari *K-Means* bisa divisualisasikan dalam bentuk diagram dua dimensi tanpa kehilangan gambaran pola pengelompokan yang penting.

Visualisasi ini sangat krusial untuk melihat secara langsung bagaimana data terdistribusi dalam berbagai klaster. Selain itu, visualisasi berbasis PCA juga mempermudah dalam menilai hasil pengelompokan, terutama untuk mengevaluasi apakah klaster yang terbentuk itu kompak, terpisah dengan jelas, atau ada tumpang tindih antara kelompok. Proses ini sangat membantu dalam memahami struktur internal data serta keabsahan pengelompokan yang dilakukan secara otomatis oleh algoritma.



Gambar 4. 4 Visualisasi PCA *Cluster*

Dari tampilan hasil pengelompokan di atas, terlihat bahwa data berhasil dibagi menjadi empat kelompok yang jelas terpisah. Setiap warna melambangkan anggota dari kelompok yang dibentuk berdasar kesamaan pola warna pada daun. Titik centroid yang ditunjukkan dengan simbol "X" berwarna merah menandakan pusat dari setiap kelompok, yang menjadi referensi utama dalam proses pembentukan kelompok oleh metode *K-Means* memperkuat temuan dari metode *Elbow* sebelumnya, yang menunjukkan bahwa nilai k terbaik adalah 4.

Dengan adanya empat kelompok yang terpisah dengan jelas, bisa memahami bahwa karakteristik warna daun mangga memiliki variasi yang cukup signifikan untuk dikelompokkan ke dalam beberapa kategori yang mungkin kelompok-kelompok ini dapat langsung dihubungkan dengan kondisi fisiologis daun, contohnya, satu kelompok mewakili daun yang sehat, sedangkan kelompok lain mencerminkan jenis penyakit tertentu seperti bercak kuning, bercak coklat, atau

tanda-tanda kekurangan nutrisi (klorosis). Proses pengelompokan ini menjadi dasar yang penting dalam sistem deteksi otomatis, karena memungkinkan pengelompokan data baru ke dalam kelompok yang telah ada tanpa harus memberikan label secara manual, sekaligus memberikan wawasan mengenai pola visual yang konsisten dalam dataset.

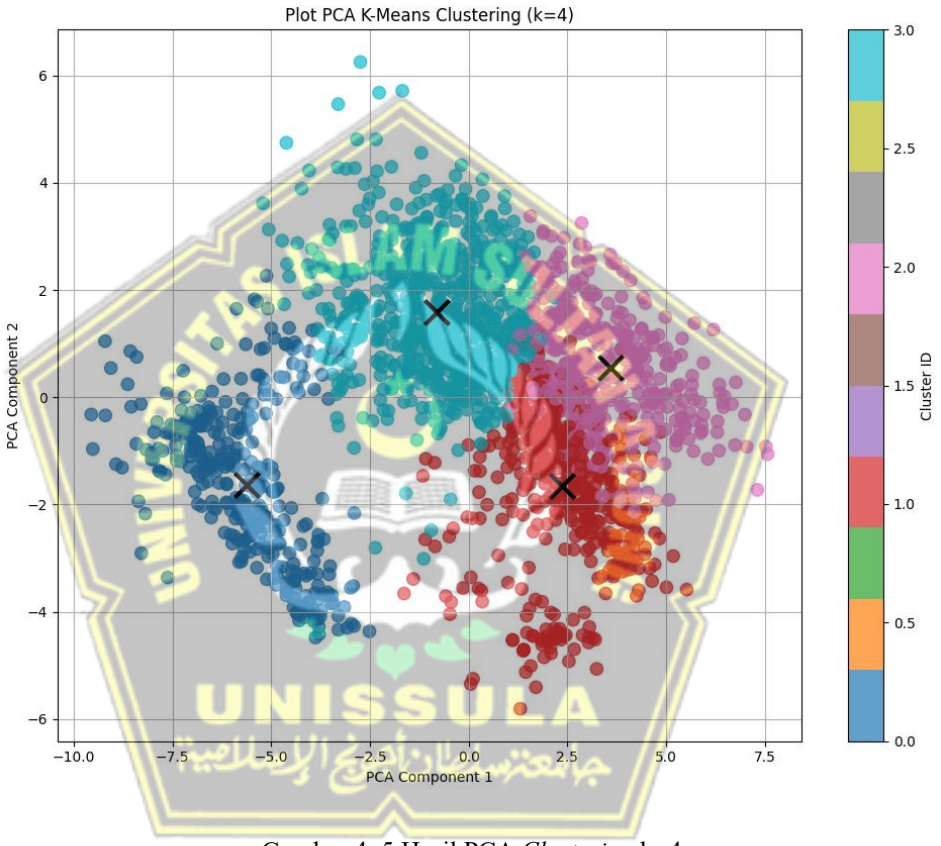
Hasil visualisasi ini dapat dimanfaatkan untuk berbagai analisis lebih lanjut, seperti melakukan interpretasi manual terhadap setiap klaster untuk mengidentifikasi jenis penyakit daun berdasarkan pola warna, atau sebagai dasar untuk pelabelan klaster sebelum digunakan dalam model klasifikasi berbasis pembelajaran terawasi. Selain itu, jika dipadukan dengan pengetahuan pakar, klaster ini juga dapat dimanfaatkan untuk menilai tingkat keparahan penyakit atau gejala visual pada daun mangga.

4.4.3 Visualisasi PCA dari *clustering* (k=4)

Setelah fitur warna dari gambar daun mangga diekstraksi, langkah berikutnya adalah mengelompokkan data tersebut menurut kesamaan nilai fiturnya. Untuk tujuan ini, algoritma *k-means clustering* diimplementasikan, yang merupakan salah satu metode *unsupervised learning* yang banyak digunakan untuk pengelompokan data tanpa label. Tujuan utamanya adalah untuk menemukan struktur alami dalam data dengan membagi data ke dalam beberapa kelompok berdasarkan jarak menuju titik pusat *(centroid)* dari setiap kelompok.

Teknik *Principal Component Analysis* (PCA) diterapkan untuk mengurangi dimensi data tersebut menjadi dua komponen utama. PCA berfungsi dengan mencari arah di mana variansi data paling besar, sehingga dua komponen utama yang dihasilkan dapat memberikan representasi visual yang cukup jelas tentang sebaran dan struktur kluster dalam data. Dengan demikian, hasil pengelompokan yang sebelumnya hanya terlihat dalam bentuk angka atau label dapat divisualisasikan dalam bidang dua dimensi, yang lebih mudah untuk dianalisis dan dipahami.

Visualisasi hasil pengelompokan ini sangat penting tidak hanya untuk menilai kinerja *k-means*, tetapi juga untuk memahami apakah pengelompokan yang terjadi sesuai dengan harapan, yakni setiap kelompok mewakili kondisi tertentu dari daun baik daun yang sehat maupun yang menunjukkan tanda-tanda penyakit. Dengan pendekatan ini, kemungkinan pola visual yang khas dari setiap jenis daun berdasarkan distribusi warnanya.



Gambar 4. 5 Hasil PCA *Clustering* k=4

Gambar di atas memperlihatkan representasi hasil *k-means clustering* pada fitur warna yang sudah direduksi dimensinya melalui *Principal Component Analysis* (PCA). Setiap titik pada gambar menggambarkan satu data citra daun, dengan warna titik tersebut menunjukkan klaster yang terbentuk berdasarkan kedekatan nilai fitur warna yang ada. Dalam visualisasi ini, digunakan nilai k=4, yang artinya algoritma *K-Means* diminta untuk membagi data menjadi 4 klaster yang berbeda.

Tanda silang merah yang terlihat di tengah-tengah kelompok menunjukkan titik pusat klaster *(centroid)* dalam ruang PCA. Titik ini dihitung sebagai rata-rata posisi semua anggota klaster dalam ruang fitur, dan digunakan sebagai referensi utama dalam proses iteratif pembentukan klaster oleh algoritma *k-means*. Jarak *Euclidean* dari setiap titik ke *centroid* menjadi dasar untuk menentukan keanggotaan titik tersebut dalam salah satu klaster.

Proses ini adalah klastering, bukan klasifikasi. Klastering berarti tidak ada label atau kategori kelas yang digunakan saat pelatihan semua kelompok terbentuk secara otomatis berdasarkan kesamaan fitur. Berbeda dengan klasifikasi, yang memerlukan data berlabel (contohnya: daun sehat, bercak hitam, bercak cokelat), klastering lebih bersifat *eksploratif* untuk menemukan pola atau struktur tersembunyi dalam data.

Hasil visualisasi ini, terlihat bahwa sebagian besar titik membentuk kelompok yang cukup jelas dengan jarak yang relatif terpisah, menunjukkan bahwa klaster yang terbentuk cukup baik dan memiliki struktur yang bisa ditafsirkan. Namun, ada juga beberapa titik yang berada di antara dua klaster, menandakan adanya potensi tumpang tindih antar kelompok. Hal ini wajar dalam analisis klastering berbasis citra, karena perbedaan warna daun bisa terjadi secara bertahap, terutama pada daun yang mengalami perubahan kondisi.

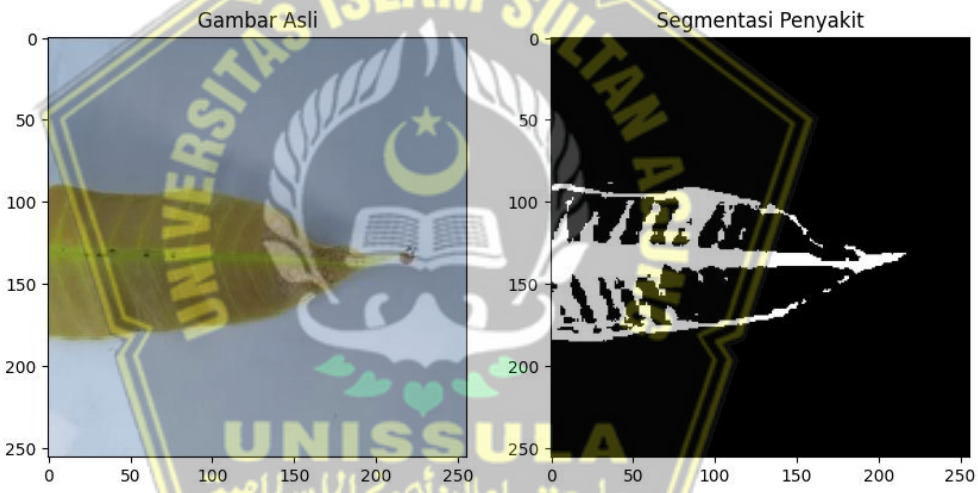
Visualisasi ini bisa menjadi langkah awal untuk interpretasi yang lebih mendalam terhadap setiap klaster. Sebagai contoh, dengan mengamati beberapa citra dari masing-masing klaster dapat menyimpulkan apakah klaster tertentu didominasi oleh daun yang sehat, atau justru oleh daun yang menunjukkan ciri-ciri penyakit tertentu.

4.5 Hasil Segmentasi

4.5.1 Segmentasi Penyakit

Segmentasi citra adalah salah satu langkah penting dalam pengolahan citra digital, terutama untuk mengidentifikasi gejala penyakit pada daun tanaman. Dalam penelitian ini, segmentasi dilakukan untuk memisahkan bagian daun mangga yang sehat dari yang berubah warna atau tekstur akibat penyakit. Metode ini

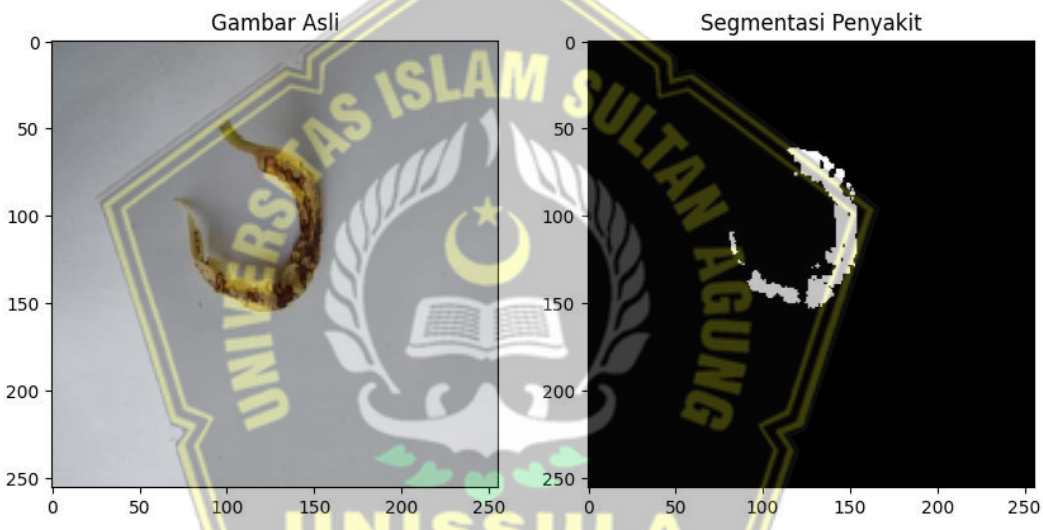
memanfaatkan teknik pengolahan citra yang berdasarkan perbedaan intensitas warna dan kontras untuk menyoroti area yang dicurigai terinfeksi. Segmentasi penyakit tidak hanya berguna untuk memvisualisasikan area yang terdampak, tetapi juga menjadi dasar untuk proses klasifikasi selanjutnya, seperti penerapan algoritma *K-Means Clustering*. Gambar di bawah ini terdapat dua sisi proses: sebelah kiri menunjukkan gambar asli daun mangga dari data uji, sedangkan sebelah kanan menunjukkan hasil segmentasi, di mana bagian putih melambangkan area yang terindikasi terkena penyakit berdasarkan perbedaan warna dan intensitas piksel dibandingkan bagian daun yang sehat. Proses ini memungkinkan sistem untuk mengenali kerusakan pada daun secara otomatis tanpa campur tangan manual, sehingga meningkatkan efisiensi dalam pemantauan kesehatan tanaman.



Gambar 4. 6 Segmentasi Penyakit Daun *Anthracnose*

Gambar diatas menunjukkan hasil segmentasi citra daun mangga yang mengalami indikasi penyakit menggunakan algoritma *k-means clustering*. Berdasarkan analisis segmentasi pada gambar tersebut, terlihat bahwa area dengan perubahan warna yang kemungkinan merupakan tanda awal atau lanjutan dari serangan penyakit berhasil dipisahkan dari bagian daun yang sehat. Area berwarna putih pada hasil segmentasi menunjukkan piksel-piksel yang teridentifikasi sebagai bagian yang tidak normal atau terinfeksi, sedangkan area hitam menunjukkan latar belakang dan bagian daun tanpa gejala. Proses ini adalah langkah penting dalam alur kerja sistem deteksi otomatis, sebab ketepatan segmentasi sangat

mempengaruhi efektivitas metode pengelompokan atau klasifikasi yang akan digunakan selanjutnya. Dengan memusatkan analisis hanya pada area yang relevan, yaitu bagian daun yang menunjukkan tanda-tanda penyakit, sistem mampu mengurangi gangguan atau data yang tidak perlu. Selain itu, hasil segmentasi mempermudah dalam pengukuran luas area yang terpengaruh, yang kemudian dapat dijadikan tolak ukur untuk menilai seberapa parah penyakit yang menyerang daun. Ini sangat berguna dalam konteks pertanian presisi, di mana keputusan untuk melakukan pengobatan atau tindakan dapat didasarkan pada data visual yang telah diolah secara sistematis dan akurat.



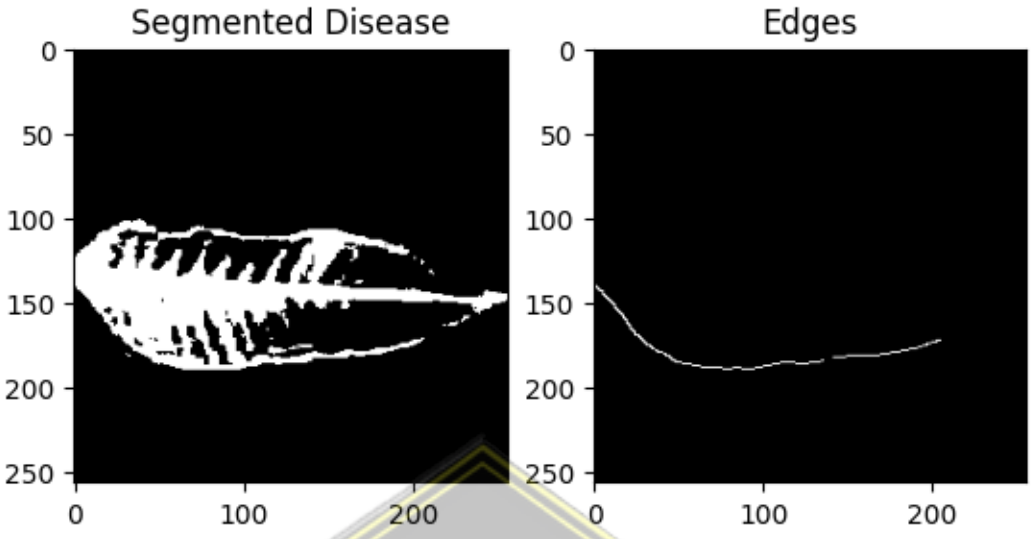
Gambar 4. 7 Segmentasi Penyakit Daun *Anthracnose*

Gambar di atas memperlihatkan dua langkah penting dalam proses mendeteksi penyakit pada daun mangga: gambar RGB asli (di sebelah kiri) dan hasil pemisahan area yang terinfeksi (di sebelah kanan). Dalam gambar asli, tampak jaringan tanaman yang mulai kering dan mengalami perubahan warna dari hijau kekuningan menjadi cokelat gelap yang secara visual menunjukkan kerusakan jaringan karena pengaruh patogen. Proses pemisahan ini kemudian melakukan ekstraksi piksel yang abnormal secara selektif, yang ditunjukkan sebagai area putih di atas latar belakang hitam. Pemisahan ini biasanya diawali dengan operasi pra-pemrosesan seperti konversi ruang warna (misalnya RGB menjadi LAB atau HSV) untuk meningkatkan kontras antara jaringan yang sehat dan yang terinfeksi, diikuti

dengan penghalusan menggunakan *Gaussian blur* untuk mengurangi *noise* latar, dan diakhiri dengan metode threshold adaptif atau pengelompokan *k-means* untuk membedakan kelas piksel. Hasil dari pemisahan ini menghasilkan objek yang terpisah dengan siluet mengikuti kontur luka, sehingga memudahkan dalam kuantifikasi seperti penghitungan area yang terinfeksi, perimeter lesi, dan rasio kerusakan terhadap total permukaan. Informasi ini tidak hanya mengonfirmasi keberadaan penyakit, tetapi juga memberikan data numerik untuk langkah yang lebih lanjut atau pemodelan tingkat keparahan. Ketepatan pemisahan sangat berpengaruh pada keabsahan analisis berikutnya: over-segmentation dapat mengakibatkan bagian yang sehat terlanjur dianggap positif palsu, sedangkan under-segmentation bisa jadi mengabaikan lesi mikro yang penting. Untuk itu, kombinasi teknik morfologi (closing/opening) dan verifikasi visual terhadap ground truth biasanya dilakukan untuk menjamin bahwa citra biner yang dihasilkan dengan tepat mencerminkan area infeksi. Dengan cara ini, tahap pemisahan ini menjadi dasar penting dalam urutan sistem otomatis untuk mendeteksi penyakit pada daun mangga menggunakan visi komputer dan kecerdasan buatan..

4.5.2 Segmentasi dengan metode *canny*

Dalam mengidentifikasi penyakit pada daun, terdapat dua langkah krusial dalam analisis citra digital, yaitu segmentasi citra dan deteksi tepi. Segmentasi berfungsi untuk memisahkan area yang terinfeksi penyakit dari bagian daun yang sehat, sedangkan deteksi tepi membantu dalam menentukan batas-batas objek yang relevan dalam citra tersebut. Salah satu metode deteksi tepi yang banyak digunakan adalah algoritma *Canny*, yang dikenal kemampuannya dalam menonjolkan kontur dengan tingkat akurasi tinggi berdasarkan perbedaan intensitas piksel. Gambar di bawah ini menunjukkan hasil dari kedua proses tersebut, yaitu segmentasi area yang terinfeksi dan deteksi tepi menggunakan metode *Canny*.



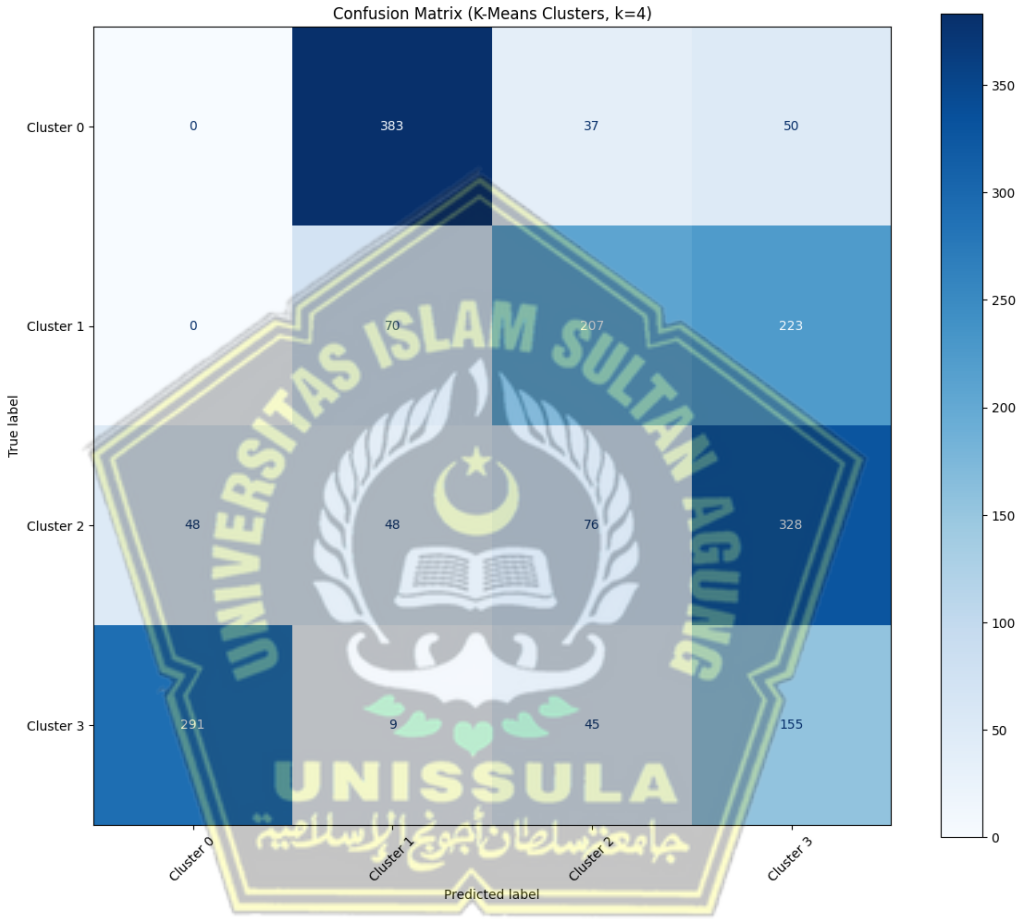
Gambar 4. 8 Segmentasi Metode *Canny*

Gambar di atas memperlihatkan di sisi kiri, terlihat hasil pemisahan area yang terjangkit penyakit pada daun mangga, di mana bagian yang terinfeksi ditandai dengan warna putih. Pemisahan ini berhasil memisahkan bagian daun yang menunjukkan tanda-tanda penyakit dari latar belakang serta jaringan daun yang masih sehat. Sementara itu, gambar di sebelah kanan menunjukkan hasil deteksi tepi melalui metode *Canny.* Teknik ini digunakan untuk menonjolkan batas luar daun dengan lebih akurat, yang sangat bermanfaat dalam analisis bentuk dan pengenalan area yang mengalami kerusakan. Dengan menggabungkan kedua metode ini pemisahan penyakit dan deteksi tepi sistem dapat memperoleh informasi visual yang lebih menyeluruh untuk mendukung proses klasifikasi atau pengelompokan selanjutnya.

4.6 Hasil

Dalam pendekatan *unsupervised learning* seperti *k-means clustering*, tujuan utamanya adalah untuk mengelompokkan data ke dalam beberapa grup berdasarkan kesamaan fitur tanpa label atau kategori yang sudah ada sebelumnya. Namun, untuk menilai seberapa baik hasil pengelompokan yang dilakukan, dapat membandingkan hasil tersebut yang sebenarnya. Salah satu metode yang sering dipakai untuk evaluasi dalam konteks ini adalah *confusion matrix*. *Confusion matrix* sangat vital

dalam mengevaluasi performa sistem klasterisasi, dari sini dapat menyimpulkan seberapa dekat hasil pengelompokan dengan kenyataan. Dalam penelitian ini, label mencakup empat kategori, yaitu *Anthracnose, Healthy, Powdery Mildew,* dan *Sooty Mould,* dengan mempertimbangkan hasil *clustering* ini, dapat menilai akurasi, serta kluster mana yang paling sering salah dalam pemetaan data.



Gambar 4. 9 *Confusion Matrix*

Berdasarkan *confusion matrix* tersebut, dapat dilihat bahwa hasil *k-means clustering* memiliki empat klaster (k = 4). Meskipun *k-means* tergolong algoritma pembelajaran *unsupervised learning*, matriks ini bermanfaat karena membandingkan label klaster yang diprediksi oleh model dengan label yang sebenarnya, sehingga dapat memberikan wawasan mengenai sejauh mana hasil segmentasi atau pengelompokan mendekati distribusi kelas yang benar. Sumbu

vertikal menunjukkan label asli data *(true label)* sedangkan sumbu horizontal merepresentasikan label klaster yang diprediksi oleh algoritma *k-means*.

Melalui *confusion matrix* ini, dapat melihat bahwa untuk cluster 0 (baris pertama), mayoritas data (383 data) berhasil ditempatkan di klaster yang tepat (prediksi cluster 1), tetapi juga terdapat kesalahan klasifikasi dengan 37 data masuk ke cluster 2 dan 50 data ke cluster 3. Situasi serupa juga terjadi di baris lainnya, di mana sejumlah data tersebar di klaster yang berbeda, menunjukkan adanya tumpang tindih fitur antar data yang menyulitkan *k-means* dalam memisahkan kelas dengan sempurna. Contohnya, pada baris keempat *(True Label Cluster 3),* sebanyak 291 data justru diklasifikasikan ke cluster 0, yang menandakan perbedaan mencolok antara klaster yang diprediksi dan label asli. Penyebabnya bisa jadi karena dua hal: pertama, data antar label memiliki kemiripan fitur yang tinggi sehingga membuat pemisahan sulit, dan kedua, karena *k-means* tidak memperhitungkan label asli dalam prosesnya, hubungan antara klaster dan label sejati menjadi tidak langsung.

Meskipun begitu, matriks ini tetap sangat penting untuk menilai kebersihan atau kualitas klaster yang terbentuk, serta mengidentifikasi potensi kesalahan yang mungkin perlu diperbaiki dengan pendekatan semi-supervised atau metode lain seperti penggabungan antara pengelompokan dan klasifikasi. Secara keseluruhan, dominasi jumlah prediksi yang akurat dalam beberapa baris (misalnya 383 pada baris pertama, 328 pada baris ketiga) menunjukkan bahwa sebagian besar data telah berhasil dikelompokkan berdasarkan pola fitur yang seragam. Visualisasi ini membantu kita dalam menilai seberapa efektif pemilihan jumlah klaster dan memahami pola distribusi dalam data dengan lebih mendalam..

Untuk mengevaluasi kinerja model dalam identifikasi berbagai jenis penyakit pada daun, digunakan *confusion matrix*. Dari *confusion matrix* tersebut, dihitunglah metrik evaluasi seperti *precision, recall,* dan *F1-score* untuk setiap kelas. Ketiga metrik ini memberikan wawasan lebih mendalam tentang kemampuan model dalam mengklasifikasikan setiap kelas dengan akurat. Hasil evaluasi ini disajikan dalam tabel berikut.

Tabel 4. 1 Matrix Evaluasi

Kelas	*Precision*	*Recall*	*F1-score*
Cluster 0	0.00	0.00	0.00
Cluster 1	0.54	0.16	0.25
Cluster 2	0.46	0.53	0.49
Cluster 3	0.38	0.29	0.33

Setelah menjalankan proses pengelompokan menggunakan algoritma *K-Means* dengan jumlah klaster empat (k = 4), evaluasi terhadap hasil pengelompokan dilakukan melalui tiga metrik utama yang umum dipakai dalam *clustering : Precision, Recall,* dan *F1-Score*. Ketiga metrik ini berguna untuk menilai seberapa baik model *clustering* merepresentasikan label asli data ketika informasi *ground truth* tersedia. Meskipun *K-Means* merupakan metode *unsupervised learning*, adanya label asli memberi kesempatan untuk membandingkan hasil klaster dengan label yang benar untuk menilai tingkat akurasi dari pengelompokan tersebut. Proses evaluasi ini sangat penting untuk mengetahui apakah klaster yang dihasilkan benar-benar mencerminkan struktur alami data yang ada.

Precision pada dasarnya menunjukkan tingkat ketepatan model dalam menentukan anggota klaster, yaitu berapa banyak data yang benar-benar berasal dari kategori tertentu dibandingkan dengan semua data yang diprediksi termasuk ke dalam klaster tersebut. Nilai *precision* yang tinggi menandakan bahwa sebagian besar data yang masuk dalam suatu klaster sesuai dengan label yang seharusnya. Sebaliknya, nilai *precision* rendah mengindikasikan bahwa banyak data dalam klaster tersebut berasal dari label yang berbeda. Dalam evaluasi, klaster 2 memiliki nilai *precision* tertinggi di antara klaster lainnya, yaitu 0.46, meskipun nilai ini masih dianggap sedang.

Recall di sisi lain, mencerminkan kemampuan model dalam menemukan semua anggota dari suatu kategori. Ini berarti bahwa *recall* mengukur seberapa banyak data dari label asli yang berhasil teridentifikasi dan dimasukkan ke dalam klaster dengan benar. Nilai *recall* yang tinggi menunjukkan bahwa sebagian besar

data dari suatu kelas tidak terabaikan dalam tahap pengelompokan. Dalam hal ini, klaster 2 juga memiliki nilai *recall* tertinggi yaitu 0.53, menunjukkan bahwa sekitar setengah dari data asli klaster 2 berhasil dikenali dan dikelompokkan dengan tepat.

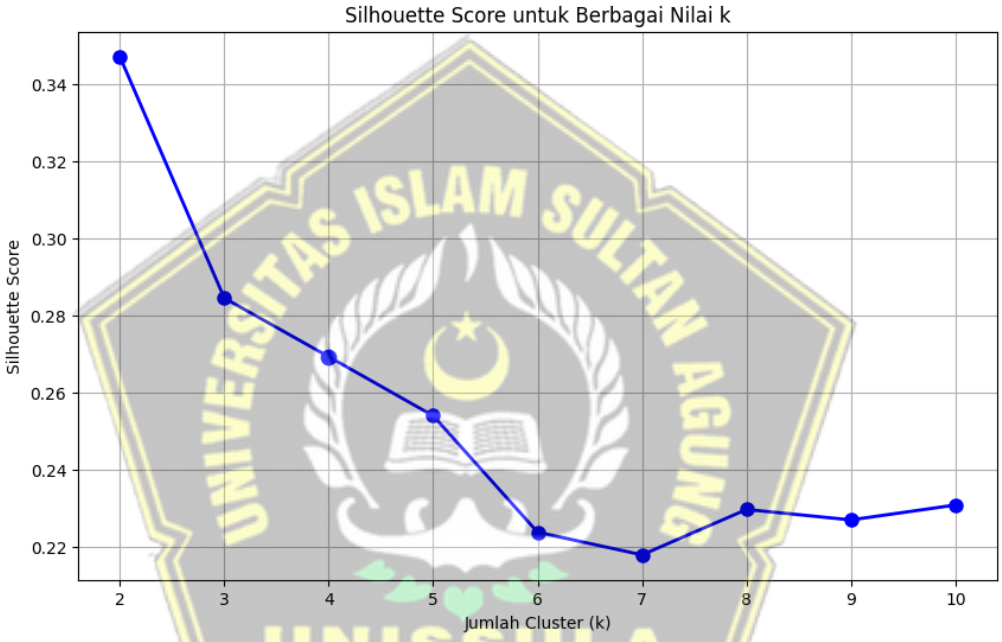
F1-Score adalah rata-rata harmonis antara *precision* dan *recall*, memberikan gambaran menyeluruh tentang keseimbangan antara kedua metrik tersebut. *F1-Score* penting karena dapat merefleksikan kinerja keseluruhan model ketika *precision* dan *recall* berada dalam kondisi yang saling bertolak belakang. Sebagai contoh, sebuah klaster dapat mempunyai nilai *recall* yang tinggi tetapi nilai *precision* yang rendah, yang berarti banyak data dari label asli berhasil dikenali tetapi juga terdapat banyak kesalahan prediksi. Dalam hasil evaluasi ini, klaster 2 kembali memperoleh angka tertinggi dalam *F1-Score* dengan nilai 0.49, sementara klaster lainnya memiliki *F1-Score* yang lebih rendah, terutama klaster 0 yang mempunyai nilai 0 untuk semua metrik. Hal ini menunjukkan bahwa klaster 0 tidak mampu mengenali data dengan label asli yang tepat, mungkin disebabkan oleh ketidaksesuaian label hasil *k-means* dengan label asli atau karena data dari klaster 0 tidak terstruktur dan tersebar ke klaster lainnya.

Berdasarkan semua metrik evaluasi yang ada, dapat disimpulkan bahwa kinerja *k-means* dalam pengelompokan data pada situasi ini masih belum berada pada tingkat optimal. Ini mungkin disebabkan oleh beberapa faktor seperti pemilihan nilai k yang kurang tepat, fitur data yang belum sepenuhnya menggambarkan variasi antar kategori, atau karena batasan dari algoritma *k-means* itu sendiri yang mengasumsikan bentuk klaster yang bulat *(spherical)* dan tidak mampu menangani tumpang tindih atau distribusi yang kompleks. Oleh karena itu, hasil ini dapat menjadi dasar untuk mengeksplorasi metode pengelompokan yang lebih *advanced* atau melakukan optimasi pada tahap *preprocessing* dan pemilihan fitur lebih lanjut agar hasil segmentasi menjadi lebih akurat dan mencerminkan struktur data yang sebenarnya.

4.7 Evaluasi Model

Selain Metode *Elbow*, penilaian jumlah klaster yang optimal juga dapat dilakukan dengan menggunakan *Silhouette Score*. *Silhouette Score* adalah ukuran

evaluasi yang menilai seberapa erat keterikatan suatu data dengan klaster yang sesuai, dibandingkan dengan klaster yang lain. Nilai ini dapat bervariasi antara -1 hingga 1, dimana nilai yang mendekati 1 menunjukkan bahwa data sangat cocok dengan klaster yang diikutinya dan sangat berbeda dari klaster lainnya. Grafik di bawah ini menunjukkan perubahan nilai *Silhouette Score* seiring dengan variasi jumlah klaster (k), mulai dari 2 hingga 10, untuk membantu menemukan nilai k terbaik dalam proses klasterisasi gambar daun mangga.



Gambar 4. 10 *Silhouette Score*

Gambar di atas memperlihatkan grafik *Silhouette Score* berkaitan dengan jumlah cluster (k) yang diterapkan dalam proses *K-Means Clustering*. *Silhouette Score* merupakan salah satu ukuran evaluasi yang krusial dalam teknik *unsupervised learning*, terutama untuk menilai kualitas hasil pengelompokan data. Ukuran ini memberikan informasi tentang sejauh mana suatu objek selaras dengan klasternya dibandingkan dengan klaster lainnya. Nilai *silhouette* berada dalam kisaran -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa objek sangat sesuai dengan klaster tersebut dan jauh dari klaster lain. Sebaliknya, nilai yang mendekati 0 menandakan bahwa objek berada di dekat batas antara dua

klaster, sedangkan nilai negatif mengindikasikan bahwa objek lebih cocok dengan klaster lain.

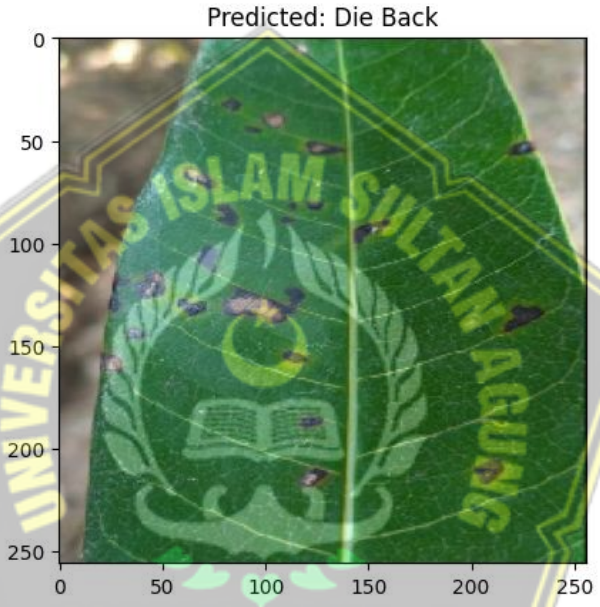
Dari grafik tersebut, terlihat bahwa skor *silhouette* tertinggi dicapai ketika jumlah klaster k = 2, dengan nilai mendekati 0,35. Ini menunjukkan bahwa saat data dibagi menjadi dua klaster, kualitas pemisahan antara klaster cukup baik, dan objek-objek dalam klaster tersebut cukup seragam. Namun, ketika jumlah klaster meningkat, nilai *silhouette* cenderung menurun. Hal ini menunjukkan bahwa penambahan klaster justru mengurangi kohesi di dalam klaster dan meningkatkan kesamaan antar klaster yang seharusnya berbeda. Sebagai contoh, pada nilai k = 6 dan k = 7, nilai silhouette turun menjadi sekitar 0,22 dan bahkan mendekati 0,21, yang menunjukkan kemungkinan model mengalami *overfitting* atau membentuk klaster yang tidak cukup berbeda satu sama lain.

Menariknya, meskipun penurunan nilai *silhouette* terlihat jelas dari k = 2 hingga k = 7, ada sedikit peningkatan dan kestabilan pada k = 8 hingga k = 10, meskipun nilainya masih rendah (sekitar 0,23). Ini bisa menjadi tanda bahwa data mulai memperlihatkan struktur baru, namun kualitas pemisahannya masih belum sebaik saat jumlah klaster lebih sedikit. Berdasarkan grafik ini, dapat disimpulkan bahwa jumlah klaster yang paling optimal berdasarkan *silhouette score* adalah k = 2, karena memberikan pemisahan antar klaster yang paling nyata dan baik. Namun, dalam praktiknya, pemilihan nilai k juga harus mempertimbangkan konteks domain, kebutuhan analisis, dan relevansi terhadap label data (jika ada). Oleh karena itu, meskipun k = 4 digunakan dalam analisis sebelumnya, dari perspektif evaluasi struktur klaster, penggunaan k = 2 sebenarnya memberikan hasil yang lebih baik dalam hal kualitas pengelompokan berdasarkan metrik silhouette.

Secara keseluruhan, grafik ini menyoroti pentingnya evaluasi metrik dalam menentukan jumlah klaster yang paling optimal, karena pemilihan nilai k yang tidak tepat dapat menyebabkan hasil *clustering* yang membingungkan atau tidak mencerminkan struktur alami data. Evaluasi *silhouette* seperti ini menjadi langkah krusial dalam proses validasi model tak terawasi agar hasil analisis data lebih berarti dan akurat.

4.8 Hasil Prediksi Citra Daun

Setelah menyelesaikan evaluasi model, langkah selanjutnya adalah mengamati performa model terhadap citra uji yang sebelumnya tidak pernah dikenali. Visualisasi ini bertujuan untuk menunjukkan secara langsung bagaimana model mengklasifikasikan kondisi daun mangga. Gambar di bawah ini menampilkan salah satu contoh hasil prediksi terhadap citra daun mangga yang menunjukkan gejala penyakit.



Gambar 4. 11 Prediksi Citra

Hasil prediksi menunjukkan bahwa model mampu mengenali pola serta gejala yang muncul pada daun tersebut, dan mengklasifikasikannya sebagai penyakit *Die Back*. Ini didasarkan pada adanya bercak atau luka berwarna coklat kehitaman yang tersebar di permukaan daun, yang merupakan salah satu ciri khas dari penyakit ini. Prediksi tersebut mencerminkan kemampuan model dalam melakukan klasifikasi berdasarkan fitur visual yang relevan, seperti tekstur dan warna area yang terinfeksi. Dengan menampilkan beberapa sampel hasil prediksi ini, terlihat bahwa model telah mempelajari karakteristik penyakit dengan cukup efektif.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma *K-Means* dalam segmentasi penyakit daun mangga yang terinfeksi penyakit menunjukkan hasil yang efektif dan efisien. Algoritma *K-Means* berhasil mengelompokkan piksel citra berdasarkan kesamaan warna dan tekstur, sehingga dapat dengan jelas membedakan antara area yang sehat dan yang terinfeksi. Hasil pengujian menunjukkan memiliki tingkat akurasi yang memadai dalam mendeteksi penyakit, yang pada gilirannya dapat membantu petani dalam pengambilan keputusan yang tepat. Dengan demikian, penelitian ini memberikan kontribusi yang signifikan terhadap deteksi penyakit tanaman yang lebih cepat dan akurat.

5.2 Saran

Meskipun penelitian ini telah menunjukkan hasil yang positif, terdapat beberapa saran untuk lebih lanjut :

1. Peningkatan Dataset: Disarankan agar penelitian selanjutnya menggunakan dataset yang lebih besar dan beragam. Hal ini diharapkan dapat meningkatkan akurasi serta generalisasi model dalam mendeteksi berbagai jenis penyakit pada daun mangga.
2. Eksplorasi Metode Lain: Penelitian lanjutan dapat menjajaki penggunaan metode segmentasi lain, seperti algoritma berbasis pembelajaran mendalam *(deep learning)*, yang mungkin menghasilkan performa yang lebih baik dibandingkan dengan K-Means.
3. Analisis *Multivariat*: Penting untuk melakukan analisis *multivariat* guna mempertimbangkan faktor-faktor lain yang dapat mempengaruhi kesehatan tanaman, seperti kondisi lingkungan dan perawatan yang diberikan. Ini dapat memberikan wawasan yang lebih mendalam dalam pengelolaan penyakit.

DAFTAR PUSTAKA

Agyztia Premana, dkk, (2020) 'Segementasi K-Means Clusteringpadacitramenggunakan Ekstraksi Fitur Warna Dan Tekstur', *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*.

Budianita, E. dkk, (2019) 'Implementasi Algoritma Canny Dan Backpropagation Untuk Mengklasifikasi Jenis Tanaman Mangga', *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)*, 11(November), pp. 13–21.

Febrinanto, F. G. dkk, (2018) 'Implementasi Algoritme K-Means Sebagai Metode Segmentasi Citra Dalam Identifikasi Penyakit Daun Jeruk', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), pp. 5375–5383. Available at: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3287.

Hidayat, D. (2022) 'Klasifikasi Jenis Mangga Berdasarkan Bentuk Dan Tekstur Daun Menggunakan Metode Convolutio Nalneural Network(Cnn) Classification of Types of Mango Based on Leave Shape and Texture Using Convolutio Nalneural Network(Cnn) Method', *Journal of Information Technology and Computer Science (INTECOMS)*, 5(1), pp. 98–103.

Kumar, S. & Singh, M. (2019) 'A novel clustering technique for efficient clustering of big data in hadoop ecosystem', *Big Data Mining and Analytics*, 2(4), pp. 240–247. doi: 10.26599/BDMA.2018.9020037.

Manalu dkk, (2023) 'Klasifikasi Penyakit Bawang Merah Melalui Citra Daun Dengan Metode K-Means', *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, 7(1), pp. 150–157. doi: 10.46880/jmika.vol7no1.pp150-157.

Mashur, M. I. & Salim, Y. (2022) 'Analisis Performa Metode Cluster K-Means pada Dataset Ocular Disease Recognition', *Indonesian Journal of Data and Science*, 3(1), pp. 35–46. doi: 10.56705/ijodas.v3i1.47.

Mulyawan dkk, (2019) 'Identifikasi Dan Tracking Objek Berbasis Image', *Identifikas Dan Tracking Objek Berbasis Image Processing Secara Real*

Time, pp. 1–5. Available at: http://repo.pens.ac.id/1324/1/Paper_TA_MBAH.pdf.

Pramudiya dkk, (2024) 'Analisis Gambar Menggunakan Metode Grayscale Dan Hsv (Hue, Saturation, Value)', *Sistem Informasi, Teknologi Informasi dan Komputer*, 14(3), pp. 174–180.

Puerwandono, E. & Maulana, I. (2023) 'Penerapan Algoritma SVM Untuk Klasifikasi Citra Daun Sirih', *INTECOMS: Journal of Information Technology and Computer Science*, 6(2), pp. 859–865. doi: 10.31539/intecoms.v6i2.7761.

Rosadi, I. dkk, (2023) 'Tingkat Keparahan Penyakit pada Daun Mangga (Mangifera indica) Menggunakan Software Imagej dan Plantix serta Kultur Bakteri pada Nutrient Agar (NA)', *BIOEDUSAINS:Jurnal Pendidikan Biologi dan Sains*, 6(2), pp. 625–633. doi: 10.31539/bioedusains.v6i2.7836.

Sinra, A. dkk, (2023) 'Optimizing Neurodegenerative Disease Classification with Canny Segmentation and Voting Classifier: An Imbalanced Dataset Study', *International Journal of Artificial Intelligence in Medical Issues*, 1(2), pp. 95–105. doi: 10.56705/ijaimi.v1i2.97.

Solikin, S. (2020) 'Deteksi Penyakit Pada Tanaman Mangga Dengan Citra Digital : Tinjauan Literatur Sistematis (SLR)', *Bina Insani Ict Journal*, 7(1), p. 63. doi: 10.51211/biict.v7i1.1336.

Suroyo, H. (2019) 'Penerapan Machine Learning dengan Aplikasi Orange Data Mining Untuk Menentukan Jenis Buah Mangga', *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 1(1), pp. 343–347. Available at: https://prosiding.seminar-id.com/index.php/sainteks/article/view/177.

Waluyo Poetro, B. S. dkk, (2024) 'Advancements in Agricultural Automation: SVM Classifier with Hu Moments for Vegetable Identification', *Indonesian Journal of Data and Science*, 5(1), pp. 15–22. doi: 10.56705/ijodas.v5i1.123.