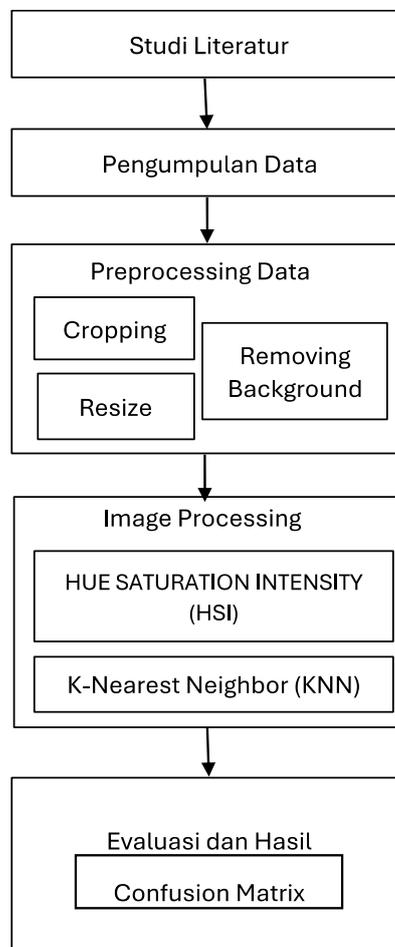


BAB III

METODOLOGI PENELITIAN

3.1. Metode Penelitian

Menurut (Sugiyono, 2010), metodologi penelitian dapat diartikan sebagai "cara-cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu." Ini mencakup strategi keseluruhan yang meliputi desain penelitian, pengumpulan data, analisis data, dan pelaporan hasil penelitian. Penelitian ini terdiri dari lima tahapan seperti pada Gambar 3.1.



Gambar 3.1 Tahapan penelitian

3.2. Studi Literatur

Studi literatur adalah proses mengumpulkan, mengevaluasi, dan menganalisis berbagai sumber informasi yang relevan dengan topik penelitian tertentu. Tujuan utama dari studi literatur adalah untuk memahami perkembangan terkini dalam bidang yang diteliti, mengidentifikasi gap atau kekosongan dalam penelitian sebelumnya, dan membangun dasar teori yang kuat untuk penelitian baru.

3.3. Pengumpulan Data

Data Pengumpulan data adalah proses pengumpulan data dimulai dengan mengambil gambar yang relevan dengan objek atau situasi yang ingin dianalisis. Dalam konteks klasifikasi kematangan buah mangga, gambar diambil dari mangga pada tingkat kematangan diantaranya yaitu tingkat kematangan mangga matang, mangga muda, dan mangga busuk. Pada penelitian ini dataset yang dikumpulkan diambil dari <https://universe.roboflow.com/ml-z0o30/mango-sjvso/dataset/3>.

3.4. Preprocessing Data

Preprocessing adalah proses menghilangkan konten yang tidak diinginkan dan untuk meningkatkan kualitas gambar agar lebih mudah dikenali. Tujuan dilakukannya *preprocessing* adalah untuk mempermudah dan memaksimalkan hasil *training* agar akurasi dari hasil akhir meningkat.

Cropping gambar adalah proses mengedit gambar dengan memotong bagian luar untuk menghilangkan area yang tidak diinginkan dan mengarahkan fokus pada bagian tertentu dari gambar. Proses ini penting dalam berbagai aplikasi pengolahan citra, karena untuk memperbaiki komposisi dan memperjelas objek utama dalam

gambar. Dalam konteks analisis citra untuk klasifikasi objek, *cropping* digunakan untuk memastikan bahwa objek yang dianalisis terpusat dan menonjol dalam *frame* gambar, mengurangi *noise* dari latar belakang yang tidak relevan. Teknik *cropping* dilakukan secara manual menggunakan perangkat lunak pengeditan gambar atau secara otomatis menggunakan algoritma komputer yang dirancang untuk mendeteksi dan memotong area penting dalam gambar. Dengan memotong gambar secara efektif, kualitas analisis dan hasil akhir dapat ditingkatkan, membuat proses pengenalan objek atau fitur lebih akurat dan efisien.

Resize pada gambar adalah proses mengubah dimensi gambar, yaitu lebar dan tinggi, tanpa mengubah isi atau proporsi utama dari gambar tersebut. Dalam konteks pembelajaran mesin, *resize* gambar sangat penting untuk memastikan bahwa semua gambar dalam dataset memiliki ukuran yang konsisten, sehingga model dapat memprosesnya dengan lebih efisien dan menghasilkan hasil yang akurat.

Removing background gambar yaitu proses menghapus atau memisahkan latar belakang dari objek utama dalam sebuah gambar. Tujuan dari teknik ini adalah untuk menonjolkan objek utama, menghilangkan elemen yang tidak diinginkan, dan menciptakan gambar dengan tampilan yang lebih bersih dan fokus. Proses ini sangat penting dalam pengolahan citra untuk keperluan analisis objek yang lebih akurat. *Removing background* dilakukan secara manual menggunakan alat pengeditan gambar seperti Adobe Illustrator. *Planning*

Planning atau perencanaan merupakan proses yang paling mendasar untuk memahami mengapa suatu sistem harus dibangun dan perencanaan pembangunan

serta penerapan arsitektur pada aplikasi yang akan dibangun, termasuk kebutuhan *software development* seperti bahasa pemrograman, *tools*, *library*, dan *resource* lainnya.

3.5. *Image Processing*

Image processing merupakan proses memanipulasi piksel dalam citra digital untuk tujuan tertentu. Mengekstraksi informasi penting dari suatu gambar dapat dilakukan dengan mengolah informasi yang ada di dalamnya. Bidang pengolahan citra digital mengacu pada pengolahan citra digital dengan menggunakan komputer digital. Pemrosesan ini bertujuan untuk mengekstraksi fitur. Penelitian ini menggunakan ekstraksi fitur warna pada pengolahan citra untuk memperoleh informasi dari citra. Ekstraksi fitur dengan memanfaatkan MATLAB untuk memperoleh nilai rata-rata HSI dari citra dan algoritma *K-Nearest Neighbor* untuk menentukan kematangan buah mangga.

1.5.1 Model Warna HSI

Model HSI adalah sistem warna yang sangat mirip dengan cara kerja mata manusia. HSI menggabungkan informasi warna dan skala abu-abu dari sebuah gambar, menjadikan ruang warna ini tampak lebih realistis dan intuitif dalam merepresentasikan warna secara alami. Model warna RGB tidak cocok untuk aplikasi pemrosesan gambar tertentu, khususnya dalam pengenalan objek yang lebih mudah mengidentifikasi objek berdasarkan perbedaan warna. Hal ini dapat dilakukan dengan menetapkan nilai ambang batas untuk rentang rona yang mencakup objek, bukan menggunakan model warna RGB.

Hue adalah sudut antara warna referensi dan vektor saturasi (S). Warna

referensi biasanya merah, tapi bisa juga warna lain. Nilai H berkisar dari 0 derajat hingga 360 derajat relatif terhadap sumbu warna merah. Sudut ini mewakili warna murni yang diencerkan oleh cahaya putih.

Saturation menggambarkan kemurnian suatu warna, seperti kuning murni atau merah murni. Parameter ini bergantung pada rentang panjang gelombang yang berkontribusi terhadap warna yang dirasakan. Sederhananya, semakin lebar rentang panjang gelombangnya, semakin kurang murni warnanya (S mendekati 0). Sebaliknya, semakin sempit rentang panjang gelombangnya, semakin murni warnanya (S mendekati 1).

Intensity adalah istilah yang cocok untuk mendeskripsikan warna beserta rona dan saturasinya. Nilai I=0 (kondisi ekstrim yang mungkin terjadi) menunjukkan warna hitam. Intensitas, yang mewakili tingkat skala abu-abu, sangat cocok untuk menafsirkan tingkat warna monokromatik. Tingkat skala abu-abu ini dapat dengan mudah diukur dan diinterpretasikan.

Untuk mengonversi ruang warna RGB ke HSI, langkah-langkah berikut diikuti:

$$H = \begin{cases} \theta & \text{if } B \leq G \\ 360 - \theta & \text{if } B > G \end{cases} \dots\dots\dots(3.1)$$

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right\} \dots\dots\dots(3.2)$$

$$S = 1 \frac{3}{(R+G+B)} [\min(R, G, B)] \dots\dots\dots(3.3)$$

$$I = \frac{1}{3}(R + G + B) \dots\dots\dots(3.4)$$

1.5.2 *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* (KNN) merupakan metode yang digunakan untuk mengklasifikasikan objek berdasarkan titik data latih terdekat pada ruang fitur. Data pelatihan diproyeksikan ke dalam ruang multidimensi, di mana setiap dimensi mewakili fitur data. KNN adalah algoritma pembelajaran terawasi, artinya algoritma ini bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola yang diketahui dengan data baru. Sebaliknya, pembelajaran tanpa pengawasan bertujuan untuk menemukan pola dalam data yang tidak memiliki label sebelumnya.

Algoritma *K-Nearest Neighbor* memiliki beberapa tahapan yaitu :

1. Menentukan parameter k .
2. Membagi dataset yaitu dengan memisahkan dataset menjadi data pelatihan dan data uji.
3. Melakukan prediksi dengan memprediksi kelas mangga matang, mangga muda dan mangga tua.
4. Menghitung akurasi prediksi

3.6. **Evaluasi Hasil**

Evaluasi hasil pada penelitian ini dilakukan dengan menggunakan metode *confusion matrix* yang biasa digunakan dalam evaluasi model klasifikasi untuk mendapatkan nilai akurasi, presisi, *recall*, dan *f1-score*.

Confusion matrix adalah alat yang digunakan dalam pembelajaran mesin dan klasifikasi statistik untuk mengevaluasi kinerja algoritma klasifikasi, tata letak tabel spesifik yang memungkinkan visualisasi kinerja suatu algoritma, biasanya

untuk pembelajaran yang diawasi di mana nilai sebenarnya diketahui. Komponen *confusion matrix* adalah matriks persegi yang membandingkan nilai prediksi dengan nilai sebenarnya. Untuk masalah klasifikasi biner, mencakup komponen-komponen berikut:

1. *True Positives* (TP) adalah jumlah kejadian yang diklasifikasikan dengan benar sebagai positif.
2. *True Negatives* (TN) adalah jumlah kejadian yang diklasifikasikan dengan benar sebagai negatif.
3. *False Positives* (FP) adalah jumlah *instance* yang salah diklasifikasikan sebagai positif (kesalahan Tipe I).
4. *False Negatives* (FN) adalah jumlah *instance* yang salah diklasifikasikan sebagai negatif (kesalahan Tipe II).

Berikut adalah struktur *confusion matrix* pada table 3.1.

Tabel 3.1 *Confusion Matrix*

	Prediksi Positif	Prediksi Negatif
Aktual Positif	TP	FN
Aktual Negatif	FP	TN

Metrik kinerja dari *confusion matrix* dengan menggunakan nilai dalam matriks konfusi, beberapa metrik penting dapat dihitung untuk mengevaluasi performa model klasifikasi dengan menggunakan :

1. Akurasi adalah rasio *instance* yang diklasifikasikan dengan benar (baik *true positif* maupun *true negative*) terhadap total *instance*.

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \dots\dots\dots(3.5)$$

2. Presisi (nilai prediktif positif) adalah rasio pengamatan positif yang diprediksi dengan benar terhadap total prediksi positif.

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \dots\dots\dots(3.6)$$

3. *Recall* (sensitivitas atau *True Positive Rate*) adalah rasio pengamatan positif yang diprediksi dengan benar terhadap semua pengamatan positif aktual.

$$\text{recall} = \frac{TP}{TP+FN} \times 100\% \dots\dots\dots(3.7)$$

4. Skor F1: Nilai rata-rata presisi dan perolehan yang harmonis, memberikan keseimbangan di antara keduanya.

$$\text{F1 Score} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \times 100\% \dots\dots\dots(3.8)$$