

### BAB III

## METODE PENELITIAN

### 3.1 Objek Penelitian

Objek penelitian ini adalah penyakit daun teh yang berfokus pada klasifikasi penyakit daun teh dengan menggunakan metode *K-Nearest Neighbors* (KNN). Dataset yang digunakan diperoleh dari *repositori Kaggle* dan terdiri dari 5.867 citra daun teh. Penelitian ini bertujuan untuk mengklasifikasikan penyakit pada daun teh berdasarkan fitur-fitur tertentu yang diambil dari citra daun, seperti warna dan tekstur. Penelitian ini bertujuan meningkatkan akurasi dan efisiensi dalam klasifikasi penyakit daun teh dibandingkan metode manual.

### 3.2 Waktu Penelitian

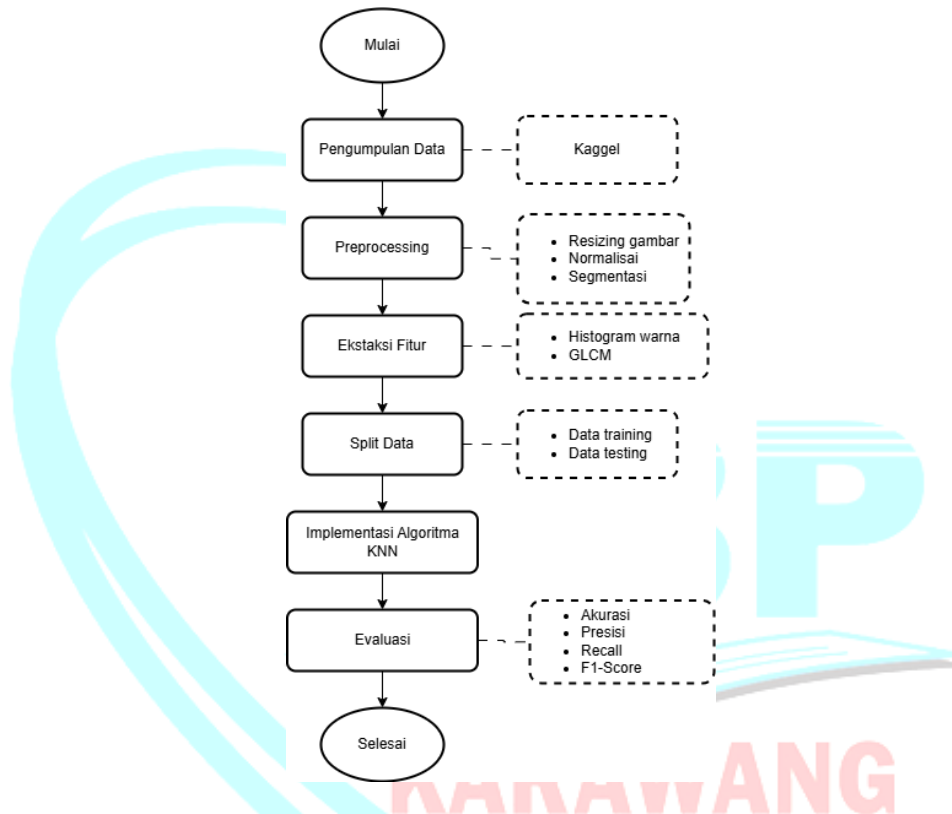
Penelitian dilakukan dengan memproses data yang telah dilakukan. Pengumpulan data berlangsung pada :

Tabel 3. 1 Waktu Penelitian

| No | Tahap penelitian       | Tahun: 2025 |         |         |         |
|----|------------------------|-------------|---------|---------|---------|
|    |                        | Bulan 1     | Bulan 2 | Bulan 3 | Bulan 4 |
| 1  | Pengumpulan Data       | ■           |         |         |         |
| 2  | preprocessing          |             | ■       |         |         |
| 3  | Ekstraksi Fitur        |             | ■       |         |         |
| 4  | Split dataset          |             |         | ■       |         |
| 5  | Implementasi Algoritma |             |         |         | ■       |
| 6  | Hasil Evaluasi         |             |         |         | ■       |

### 3.3 Prosedur Penelitian

Untuk memperoleh hasil yang diharapkan dalam penelitian, diperlukan langkah-langkah penelitian yang terstruktur. Berikut alur prosedur yang digunakan dapat dilihat pada Gambar 3.1



Gambar 3. 1 Prosedur Penelitian

Penjelasan Alur :

1. Pengumpulan *Dataset*

Pengumpulan dataset gambar daun teh dari *repositori Kaggle* yang mencakup berbagai kondisi penyakit.

2. *Preprocessing*

Data diproses lebih lanjut untuk meningkatkan kualitasnya tahapan ini mencakup :

a. *Resizing* gambar

*Resizing* memastikan semua gambar memiliki ukuran yang sama sebelum dimasukkan ke dalam model

b. *Normalisasi*

*Normalisasi* dilakukan untuk menyamakan skala nilai piksel dalam gambar agar lebih mudah diolah oleh algoritma.

c. *Segmentasi*

*Segmentasi* dilakukan untuk memisahkan objek dengan latar belakang untuk mempermudah dan memfokuskan analisis pada bagian penting citra.

3. *Ekstraksi Fitur*

Sebelum melatih model, fitur penting dari gambar diambil menggunakan metode seperti GLCM (*Gray-Level Co-occurrence Matrix*) dan LBP (*Local Binary Pattern*) untuk tekstur dan HSV (*Hue, Saturation, Value*) untuk representasi warna.

4. *Split Data*

Dataset dibagi menjadi dua bagian yaitu data latih untuk melatih model dan data uji untuk mengevaluasi performa model. Pembagian dilakukan dengan proporsi 80:20.

5. *Implementasi Algoritma K-Nearest Neighbor (KNN)*

Algoritma KNN diterapkan pada data latih untuk membangun model klasifikasi. Algoritma ini menghitung jarak antara data uji dan data latih berdasarkan fitur yang diekstraksi.

a. Menentukan parameter  $K$

b. Menghitung jarak antara data uji dan data latih menggunakan metrik jarak *Euclidean*.

- c. Menentukan nilai  $K$  sebagai jumlah tetangga terdekat.
- d. Mengurutkan jarak dan memilih  $K$  tetangga terdekat.

## 6. Hasil Evaluasi

Mengevaluasi performa model berdasarkan metrik seperti akurasi, *presisi*, *recall* dan *F1-score* untuk memastikan klasifikasi penyakit teh berjalan dengan baik. Untuk menghitung evaluasinya sebagai berikut :

### a. Akurasi

Berikut merupakan rumus untuk melakukan perhitungan akurasi pada evaluasi model:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (1)$$

Keterangan :

TP (*True Positive*): Jumlah sampel *positif* yang berhasil dikenal sebagai *positif*.

TN (*True Negative*): Jumlah sampel *negatif* yang berhasil diklasifikasi dengan benar sebagai *negatif*.

FP (*False Positive*): Jumlah data *negatif* yang salah diklasifikasikan sebagai *positif*.

FN (*False Negative*): Jumlah data *positif* yang salah diklasifikasikan sebagai *negatif*.

### b. Presisi

Menilai sejauh mana hasil prediksi benar terhadap seluruh prediksi *positif* yang dihasilkan oleh model. Rumus untuk mengukur *presisi* yaitu:

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100$$

Keterangan :

TP (*True Positive*): prediksi benar pada kelas *positif*.

FP (*False Positive*): prediksi salah pada kelas *positif*.

c. *Recall*

Mengukur sejauh mana model berhasil menemukan setiap entri positif dalam dataset. Rumus untuk mengukur *recall* yaitu:

$$Recall = \frac{TP}{TP+FP} \times 100 \quad (3)$$

Keterangan :

TP : *True positive*

FN : *False negative*

d. *F1-Score*

Menggabungkan *presisi* dan *recall* dalam satu metrik harmonis, terutama berguna ketika ada ketidak seimbangan data. Rumus untuk mengukur *F1-Score* adalah sebagai berikut:

$$F1-Score = 2 \times \frac{Presisi \times Recall}{Presisi + recall} \quad (4)$$