



1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.  
Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.  
2. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.  
Plagiarisme juga dilarang dan dapat dikenakan sanksi.  
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

## BAB III

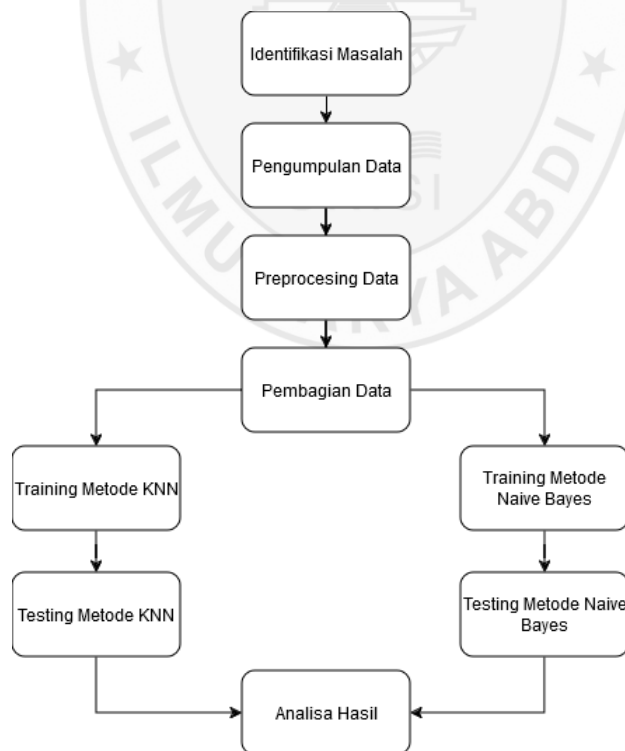
### METODE PENELITIAN

Pada bab ini menjelaskan tentang bagaimana metode atau cara yang dilakukan dalam melakukan perbandingan metode klasifikasi *K-Nearest Neighbor* dan *Naïve Bayes* terhadap data tingkat kematangan kelapa di Tembilahan dan menjelaskan teknik pengolahan data yang dilakukan sehingga mampu disesuaikan dengan algoritma klasifikasi yang digunakan.

#### 3.1. Kerangka Penelitian

Ada beberapa tahapan kegiatan dalam proses perbandingan metode klasifikasi, dalam kerangka penelitian ini mengacu pada diagram yang menjelaskan tata cara dalam melakukan perbandingan metode klasifikasi.

Tahapan tersebut disajikan pada kerangka penelitian seperti berikut:



Gambar 3.1 Kerangka Penelitian



### 3.2. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif yang bertujuan untuk menganalisis performa metode klasifikasi *k-Nearest Neighbor* (KNN) dan *Naive Bayes* pada data tingkat kematangan kelapa. Pendekatan ini digunakan untuk memperoleh hasil yang terukur secara objektif, dengan fokus pada perbandingan akurasi, presisi, dan efisiensi kedua metode.

### 3.3. Lokasi dan Waktu Penelitian

Penelitian dilakukan dengan menggunakan data tingkat kematangan kelapa yang diambil dari Kabupaten Tembilahan, yang merupakan salah satu daerah penghasil kelapa terbesar di Indonesia. Perkiraan waktu penelitian dimulai dari pengumpulan data pada bulan Januari hingga pengolahan dan analisis data selesai pada bulan Mei 2024.

### 3.4. Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data primer berupa tingkat kematangan kelapa yang diperoleh langsung dari petani dan pelaku industri kelapa di Tembilahan. Data mencakup parameter-parameter seperti warna kulit, berat buah, diameter lingkaran buah, dan kondisi tekstur buah, yang merupakan indikator tingkat kematangan kelapa. Setiap buah ditimbang beratnya dan diukur setiap diameter buah tersebut satu persatu. Untuk tekstur buah di Pada warna kulit data kematangan kelapa ini dikonversikan menjadi RGB dengan aplikasi dari android yaitu color converter.



### 3.5. Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui observasi langsung di lapangan dan wawancara terstruktur dengan petani kelapa di Tembilahan. Observasi digunakan untuk mengidentifikasi parameter visual seperti warna, tekstur, dan bentuk kelapa, sedangkan wawancara bertujuan untuk mendapatkan informasi tambahan terkait klasifikasi tingkat kematangan kelapa berdasarkan pengalaman petani. Data yang diperoleh kemudian dicatat dan disusun dalam format yang siap untuk diolah lebih lanjut.

#### 3.5.1. Observasi Lapangan:

Mengamati langsung proses panen dan pengolahan kelapa di lokasi penelitian untuk memastikan relevansi data yang diambil.

#### 3.5.2. Wawancara Tidak Terstruktur:

Menggunakan tanya jawab langsung terhadap petani setempat yang tidak memiliki format atau pernyataan yang tetap untuk memperoleh informasi dari petani dan pelaku industri terkait klasifikasi kematangan kelapa.

### 3.6. Teknik Pengolahan Data

Data yang telah dikumpulkan melalui observasi dan wawancara diproses dalam beberapa tahap. Tahap pertama adalah preprocessing data dan tahap kedua adalah pembagian data. Berikut dua tahapan dalam pengolahan data tersebut[17]:

#### 3.6.1. Preprocessing Data

Data mentah diolah untuk menghilangkan anomali seperti nilai yang hilang atau tidak relevan. Data juga di-normalisasi agar sesuai dengan kebutuhan algoritma klasifikasi.



### 3.6.2. Pembagian Data:

Data dibagi menjadi dua bagian, yaitu training set (80%) untuk melatih model dan test set (20%) untuk evaluasi performa model.

### 3.7. Metode Klasifikasi

Metode klasifikasi yang digunakan dalam penelitian ini melibatkan dua pendekatan utama, yaitu *K-Nearest Neighbor* (KNN) dan *Naive Bayes*.

#### 3.7.1. *K- Nearest Neighbor* (KNN)

Metode KNN adalah salah satu algoritma klasifikasi yang berbasis pada prinsip jarak antara data. KNN bekerja dengan cara membandingkan data yang akan diklasifikasikan dengan data dalam training set berdasarkan kedekatan jarak Euclidean atau metrik lainnya. Algoritma ini menentukan kategori data baru berdasarkan mayoritas kategori dari  $k$  tetangga terdekatnya, di mana nilai  $k$  adalah parameter yang ditentukan sebelumnya. Dalam penelitian ini, KNN diterapkan untuk mengukur efektivitasnya dalam mengenali pola-pola tingkat kematangan kelapa. Variasi nilai  $k$  akan diuji untuk menentukan nilai optimal yang memberikan akurasi terbaik. Berikut ini rumus hitung bobot jarak dalam penerapan menggunakan Jarak *Euclidean* [14]:

$$d_{(xy)} = \sqrt{\sum_{j=1}^m (X_i - Y_i)^2}$$

Ket:

$d$  : Nilai jarak (kedekatan)

$x$  : Data latih

$y$  : Data uji

$n$  : Total atribut 1 . . . n

I : Atribut nilai antara 1 . . . n

### 3.7.2. *Naïve Bayes*

Sementara itu, *Naive Bayes* adalah algoritma klasifikasi berbasis probabilistik yang menggunakan Teorema Bayes sebagai dasar perhitungan. Algoritma ini mengasumsikan bahwa setiap fitur data bersifat independen satu sama lain (naive assumption), meskipun dalam kenyataan sering kali tidak sepenuhnya demikian. Keunggulan utama *Naive Bayes* adalah kesederhanaannya serta efisiensi komputasi, sehingga sangat cocok untuk data dengan jumlah atribut yang besar. Dalam penelitian ini, *Naive Bayes* digunakan untuk mengklasifikasikan data tingkat kematangan kelapa berdasarkan distribusi probabilitas dari setiap fitur, seperti warna kulit, berat buah, dan kondisi daging kelapa. Adapun rumus dalam menghitung *Naïve bayes* seperti persamaan berikut[15]:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Ket:

X : Data class belum ditemukan

H : Hipotesis data

P(H|X): Probabilitas hipotesis berdasarkan kondisi x

P(X|H): Probabilitas berdasarkan kondisi hipotesis

P(H) : Probabilitas hipotesis H

P(X) : Probabilitas X

Kedua metode ini akan dievaluasi performanya berdasarkan kriteria akurasi, presisi, dan efisiensi. Analisis komparatif ini bertujuan untuk menentukan metode





mana yang paling optimal dalam mengklasifikasikan data tingkat kematangan kelapa di Tembilahan. Kombinasi kedua metode ini memberikan pandangan yang komprehensif tentang kemampuan teknologi kecerdasan buatan dalam mendukung sektor agrikultur.

### 3.8. Metode Perbandingan

Jenis metode perbandingan yang digunakan dalam penelitian ini adalah pendekatan komparatif kuantitatif. Metode ini digunakan untuk membandingkan performa dua algoritma klasifikasi berdasarkan hasil evaluasi numerik dari pengujian yang sama. Analisis dilakukan dengan cara mengamati dan membandingkan hasil perhitungan metrik evaluasi utama (akurasi, *precision*, *recall*, *F1-score*, dan *kappa*) yang diperoleh dari masing-masing model, sehingga memungkinkan untuk menilai metode mana yang paling sesuai dengan karakteristik data tingkat kematangan kelapa[20].

### 3.9. Kriteria Evaluasi

Kinerja metode KNN dan *Naive Bayes* dievaluasi menggunakan enam kriteria utama yaitu[10][17]:

1. Akurasi: Mengukur persentase data yang diklasifikasikan dengan benar.

Berikut adalah rumus yang digunakan dalam menentukan akurasi dari metode yang digunakan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Ket:

TP : Jumlah positif benar

FP : Jumlah positif palsu



FN : Jumlah negatif palsu

2. Presisi dan Recall: Mengukur kemampuan model dalam memprediksi data positif dengan benar serta sensitivitas model terhadap data sebenarnya. Untuk menentukan presisi dan recall pada suatu klasifikasi digunakan rumus sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

Ket:

TP : Jumlah positif benar

FP : Jumlah positif palsu

FN : Jumlah negatif palsu

- 4 F1 Score: Rata-rata harmonik dari recall dan precision. Berikut rumus yang digunakan:

$$F_1 = 2 \cdot \frac{Precision \cdot recall}{Precision + recall}$$

- 5 Cohen's Kappa: Mengukur kesepakatan antara dua anotator (atau model dan ground truth), dikoreksi terhadap kesepakatan yang terjadi secara kebetulan. Berikut rumus yang digunakan untuk menghitung Cohen's Kappa:

$$k = \frac{P_o - P_e}{1 - P_e}$$

Ket:

Po : **Observed accuracy** (jumlah prediksi benar dibagi total)



Pe : **Expected accuracy by chance**, dihitung dari total proporsi prediksi dan aktual di masing-masing kelas

## 5.1. Alat dan Perangkat yang Digunakan

Penelitian ini menggunakan kombinasi perangkat keras dan perangkat lunak untuk mendukung proses pengolahan dan analisis data. Perangkat keras yang digunakan adalah laptop atau PC dengan spesifikasi minimum prosesor Intel Core i5, RAM 4GB, dan penyimpanan SSD untuk memastikan kelancaran proses analisis data yang memerlukan daya komputasi cukup tinggi. Dari sisi perangkat lunak, penelitian ini memanfaatkan beberapa aplikasi, yaitu RapidMiner dan Tableau. RapidMiner digunakan untuk proses klasifikasi data menggunakan metode KNN dan *Naïve Bayes* dengan pendekatan antarmuka visual yang intuitif, mempermudah eksperimen parameter. Tableau digunakan untuk memvisualisasikan hasil analisis agar lebih mudah dipahami oleh pembaca dan pemangku kepentingan.



1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

2. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia. Plagiarisme juga dilarang dan dapat dikenakan sanksi.

3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.