

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Penelitian Terkait**

Dalam penelitian ini, penulis sedikit banyak mengambil referensi dari penelitian-penelitian sebelumnya yang berkaitan dengan topik pada penelitian ini, Berikut akan peneliti jabarkan beberapa peneliti terkait yang pernah dibuat sebelumnya.:

Penelitian oleh (Saputra & Andriyani, 2025). sistem klasifikasi gambar untuk buah-buahan segar dan busuk khususnya, apel, pisang, dan jeruk dikembangkan menggunakan algoritma Naïve Bayes dengan ekstraksi fitur melalui Histogram Oriented Gradients (HOG). Dataset yang digunakan terdiri dari 13.599 gambar yang dibagi menjadi data pelatihan (80%) dan data pengujian (20%). Prosesnya dimulai dengan prapemrosesan gambar menjadi skala abu-abu, mengekstraksi fitur HOG, dan melatih model Gaussian Naïve Bayes. Hasil klasifikasi menunjukkan akurasi 87%, dengan presisi tertinggi pada kelas apel segar (0,9792) dan tingkat recall tertinggi pada kelas apel busuk (0,9843). Meskipun terdapat beberapa kesalahan klasifikasi, terutama dalam kasus jeruk busuk, sistem ini menunjukkan potensi besar untuk identifikasi kualitas buah berbasis gambar. Disarankan untuk menggunakan fitur tambahan seperti warna RGB dan mempertimbangkan model klasifikasi yang lebih kompleks seperti SVM atau CNN untuk mendapatkan hasil yang lebih optimal.

Penelitian (Afriansyah et al. 2023) ini bertujuan untuk mengelompokkan tingkat kemanisan buah apel (manis, sedang, dan asam) berdasarkan atribut warna

RGB dengan memanfaatkan algoritma Naïve Bayes. Proses dimulai dengan pengumpulan 150 gambar apel, ekstraksi fitur warna (nilai R, G, B, rata-rata, dan deviasi standar), serta pembagian data melalui 3-fold cross validation untuk pelatihan dan pengujian. Klasifikasi menunjukkan bahwa metode Naïve Bayes dapat mengenali tingkat kemanisan apel dengan rata-rata akurasi 88,88%, di mana kelas manis mencapai akurasi 100%, sementara kelas sedang 86,66%, dan kelas asam 80%. Walaupun ada kekeliruan dalam pengklasifikasian pada kategori sedang dan asam, studi ini menunjukkan keberhasilan metode Naïve Bayes dalam mendeteksi pola warna gambar apel sebagai indikator tingkat kemanisan dengan cara yang sederhana dan efisien.

Penelitian oleh (Haba & Pelangi, 2020). ini bertujuan untuk mengkategorikan tingkat kematangan buah jeruk keprok (matang dan tidak matang) dengan memanfaatkan metode ekstraksi fitur tekstur Gray Level Co-occurrence Matrix (GLCM) dan algoritma klasifikasi Naïve Bayes. Proses dimulai dengan pengumpulan 130 gambar jeruk, diubah menjadi grayscale, diekstraksi nilai fitur GLCM seperti kontras, korelasi, energi, dan homogenitas, kemudian diklasifikasikan dengan Naïve Bayes. Dataset terdiri dari 110 gambar untuk pelatihan dan 20 gambar untuk pengujian. Hasil uji coba menunjukkan bahwa sistem klasifikasi ini mencapai akurasi 82%, yang membuktikan bahwa metode ini efektif serta efisien dalam mendukung petani dalam proses pengelompokan buah secara digital dan otomatis.

Penelitian oleh (Fatmiatun, 2023) bertujuan untuk mengembangkan sistem klasifikasi tingkat kematangan buah pisang kepok dengan memanfaatkan citra

digital dan algoritma Naïve Bayes. Penelitian ini berfokus pada kebutuhan petani untuk mengelompokkan buah pisang dengan efisien menurut tingkat kematangan, yang selama ini dilakukan secara manual dan tidak efektif. Sistem ini menggunakan citra warna HSV sebagai fitur utama, dengan total 80 data citra yang terbagi menjadi 60 data pelatihan dan 20 data pengujian. Proses dimulai dari pengambilan gambar, pra-pemrosesan (konversi RGB ke grayscale), segmentasi, ekstraksi fitur warna HSV, dan diakhiri dengan klasifikasi menggunakan algoritma Naïve Bayes dalam lingkungan MATLAB. Hasil percobaan menunjukkan bahwa sistem ini dapat mencapai akurasi klasifikasi hingga 90%, yang menunjukkan keefektifan metode yang diterapkan dalam membantu petani untuk mengkategorikan buah pisang berdasarkan tingkat kematangannya dengan lebih cepat dan tepat.

Penelitian oleh (Syuja, 2025) bertujuan untuk mengembangkan sistem klasifikasi otomatis tingkat kematangan buah apel emas dengan membandingkan kinerja metode Gray Level Co-Occurrence Matrix (GLCM) dan Convolutional Neural Network (CNN). Studi ini memanfaatkan citra apel emas yang dikelompokkan menjadi tiga kelas (Apple Golden 1, 2, dan 3), melalui tahapan preprocessing, augmentasi data, ekstraksi fitur GLCM (seperti kontras, energi, homogenitas, entropi), dan pelatihan model CNN. Temuan penelitian mengindikasikan bahwa model CNN menunjukkan hasil klasifikasi yang jauh lebih baik dibandingkan GLCM, dengan nilai akurasi, presisi, recall, dan F1-score mencapai angka sempurna (1.00), sedangkan model GLCM hanya sekitar 0.67. Penelitian ini menyatakan bahwa penggabungan GLCM dan CNN dapat

meningkatkan akurasi klasifikasi, tetapi CNN sendiri lebih efisien dalam mengenali citra visual yang rumit dari apel emas.

Berdasarkan kajian terdahulu menunjukkan bahwa sebagian besar penelitian tentang klasifikasi kualitas buah, khususnya apel, hanya menggunakan fitur citra tertentu seperti warna atau tekstur. Beberapa penelitian sebelumnya juga membatasi metode klasifikasi pada algoritma tertentu tanpa mengoptimalkan proses ekstraksi fitur yang lebih beragam. Penelitian ini membuat vektor fitur yang lebih komprehensif dengan menggabungkan fitur warna (Rmean, Gmean, Bmean), fitur tekstur (Contrast, Correlation, Energy, Homogeneity), dan fitur statistik citra (Entropy, Smoothness, Skewness, dan Intensitas) yang diperoleh dari analisis GLCM. Selain itu, penelitian ini menerapkan algoritma Naive Bayes dalam antarmuka GUI MATLAB yang terintegrasi. Antarmuka ini memungkinkan penggunaan proses interaktif untuk pelatihan model, pengujian, evaluasi, dan ekspor hasil. Oleh karena itu, penelitian ini tidak hanya berfokus pada teknik klasifikasi, tetapi juga pada kemudahan penggunaan, dokumentasi hasil, dan analisis akurasi. Semua ini dapat diterapkan secara langsung pada penentuan kualitas buah apel Malang Manalagi.

## **2.2 Pengolahan Citra Digital**

Pemrosesan citra digital adalah bidang ilmu komputer yang mempelajari penggunaan komputer dalam memanipulasi citra digital. Keluarannya dapat berupa informasi yang diekstrak dari suatu citra atau peningkatan kualitas visual citra itu sendiri. Citra digital adalah representasi berbasis data diskrit yang menyimpan

tampilan suatu objek. itra digital terdiri atas array dua dimensi dari piksel-piksel, di mana setiap piksel memiliki nilai intensitas yang spesifik.

Dalam penelitian ini, pemrosesan citra buah apel Malang Manalagi dilaksanakan dengan menggunakan antarmuka grafis berbasis MATLAB (GUI). Proses pemrosesan dilaksanakan melalui beberapa langkah berikut:

#### 2.2.1 Akuisisi Citra

Citra apel diperoleh dalam format RGB (.jpg/.png) dan dikelompokkan ke dalam dua kategori: segar dan busuk. Gambar-gambar ini dimasukkan ke dalam folder data\_training dan data\_uji sesuai dengan tujuannya.

#### 2.2.2 Pra-pemrosesan (Preprocessing)

Setiap citra diubah ukurannya menjadi 100x100 piksel menggunakan fungsi `imresize`, sehingga fitur dapat diekstraksi secara konsisten tanpa dipengaruhi oleh variasi ukuran citra. Selanjutnya, citra dikonversi ke format grayscale menggunakan fungsi `rgb2gray`, agar fitur tekstur dan statistik seperti GLCM, entropy, dan smoothness dapat dihitung.

#### 2.2.3 Ekstraksi Fitur

Setelah tahap preprocessing, fitur diambil dengan memanfaatkan fungsi `ekstrakFitur` yang telah dibuat dalam kode MATLAB. Fitur yang diambil mencakup fitur warna (Rmean, Gmean, Bmean), tekstur (Kontras, Korelasi, Energi, Homogenitas dari GLCM), serta statistik (Entropi, Kelancaran, Kemiringan, dan Intensitas).

#### 2.2.4 Klasifikasi

Fitur-fitur yang diekstraksi dipakai sebagai input untuk model klasifikasi Naive Bayes. Model ini dilatih memakai fungsi `fitnb` dan disimpan dalam berkas `model_nb.mat`. Prediksi pada citra uji dilaksanakan dengan fungsi `predict`, dan hasilnya dibandingkan dengan label yang sebenarnya untuk mengevaluasi akurasi klasifikasi.

#### 2.2.5 Evaluasi

Selain klasifikasi langsung, sistem menyediakan validasi silang (cross validation) 5-fold untuk mengevaluasi akurasi model secara keseluruhan dan mencegah terjadinya overfitting. Fitur ini dapat diakses melalui tombol “Crossval Akurasi “ di antarmuka pengguna grafis.

#### 2.2.6 Antarmuka Pengguna

Menawarkan kemudahan dalam interaksi lewat tombol-tombol seperti Latih Model, Jelajahi Gambar, Uji Semua, dan Atur Ulang. Pengguna dapat segera mengamati citra asli dan grayscale, serta tabel hasil klasifikasi secara langsung.

Dengan cara ini, pemrosesan gambar secara digital tidak hanya berlangsung dengan cara yang efisien dan sistematis, tetapi juga dapat digunakan oleh orang-orang yang bukan ahli teknologi melalui bantuan antarmuka visual. Sistem ini sangat mendukung penerapan nyata di sektor pertanian, terutama dalam proses penyortiran dan penentuan kualitas apel secara otomatis dan tanpa bias.

## 2.3 Ekstraksi Fitur

Ekstraksi fitur merupakan langkah untuk mendapatkan informasi penting dari gambar yang bermanfaat dalam membedakan satu objek dari yang lainnya. Tahapan ini sangat krusial dalam sistem pengenalan pola atau pengklasifikasian gambar. Tanpa fitur yang tepat dan bermanfaat, Tingkat akurasi sistem klasifikasi akan menurun. Fitur dapat berupa karakteristik warna, tekstur, bentuk, atau data statistic yang terdapat dalam gambar. Dalam penelitian ini, fitur dikelompokkan menjadi tiga jenis, yaitu fitur warna, fitur tekstur, dan fitur statistic.

### 2.3.1 Fitur Warna

Fitur warna dimanfaatkan untuk menggambarkan informasi visual yang terkait dengan susunan warna dalam sebuah gambar. Warna merupakan salah satu aspek krusial dalam identifikasi objek karena dapat berfungsi sebagai petunjuk mengenai kualitas atau keadaan suatu objek. Dalam gambar RGB, warna diwakili dalam tiga saluran yakni merah, hijau dan biru. Dalam studi ini, fitur warna yang diambil adalah rata-rata dari setiap saluran warna yaitu:

- a. Rmean: Rata-rata nilai intensitas warna merah.
- b. Gmean: Rata-rata nilai intensitas warna hijau.
- c. Bmean: Rata-rata nilai intensitas warna biru.

Warna dari apel yang masih baru umumnya lebih terang dan konsisten, sedangkan apel yang sudah busuk biasanya menunjukkan warna yang lebih gelap terang, kurang bercahaya, atau tidak konsisten. Maka dari itu, rata-rata nilai intensitas warna bisa dijadikan patokan untuk menilai mutu apel. Aspek warna juga

memiliki keuntungan karena mudah untuk diukur dan sangat memberikan informasi untuk objek yang memiliki ciri warna tertentu seperti apel.

### **2.3.2 Fitur Tekstur (GLCM)**

Tekstur adalah sifat permukaan suatu objek yang menunjukkan bagaimana warna atau intensitas piksel terdistribusi dalam gambar. Informasi Tekstur bisa menggambarkan tekstur halus, kasar, atau pola tertentu dari permukaan tersebut. Salah satu cara yang sering digunakan untuk mengambil informasi tekstur adalah dengan menggunakan metode Gray Level Co-occurrence Matrix (GLCM). Metode ini bekerja dengan menghitung seberapa sering pasangan nilai warna tertentu muncul dalam gambar, dengan jarak dan arah tertentu. Dari hasil perhitungan ini, kita bisa mengetahui berbagai parameter yang menjelaskan karakteristik tekstur tersebut.

- a. Contrast: Mengukur tingkat perbedaan kontras lokal dalam citra. Nilai kontras tinggi menunjukkan adanya variasi intensitas yang besar.
- b. Correlation: Menunjukkan sejauh mana piksel satu berhubungan secara statistik dengan piksel lainnya.
- c. Energy: Mengukur keseragaman atau kehalusan tekstur. Nilai energy tinggi menunjukkan tekstur yang seragam.
- d. Homogeneity: Menunjukkan seberapa dekat nilai piksel terhadap diagonal matriks GLCM. Nilai tinggi menunjukkan tekstur yang homogen.

Fitur tekstur penting karena memberikan informasi tambahan yang tidak bisa didapat hanya dari fitur warna saja. Apel yang sudah busuk biasanya memiliki



tekstur yang lebih kasar, sementara apel yang masih segar biasanya memiliki permukaan yang lembut dan rata.

### 2.3.3 Fitur statistik

Fitur statistik memberikan gambaran umum mengenai bagaimana nilai intensitas terdistribusi dalam sebuah gambar. Fitur ini digunakan untuk memperbaiki cara gambar terlihat dari segi warna dan teksturnya. Beberapa fitur statistik yang digunakan dalam penelitian ini antara lain:

#### a. Entropi

Mengukur seberapa acak atau tidak teratur nilai intensitas dalam gambar. Jika entropi tinggi, maka gambar tersebut memiliki banyak informasi yang tidak terduga atau cukup rumit.

#### b. Smoothness

Menunjukkan seberapa halus permukaan objek dalam gambar. yang lebih halus biasanya memiliki nilai smoothness yang lebih besar.

#### c. Skewness

Mengukur seberapa tidak simetris distribusi nilai intensitas. Nilai skewness bisa menunjukkan apakah distribusi lebih condong ke arah kiri atau kanan.

#### d. Intensitas

Merupakan rata-rata nilai intensitas dari setiap piksel dalam gambar grayscale, yang mencerminkan seberapa terang atau gelap gambar tersebut secara umum.

Dengan menggabungkan fitur warna, tekstur, dan statistik, sistem klasifikasi bisa

memahami lebih dalam tentang ciri-ciri visual objek, sehingga hasil klasifikasi menjadi lebih tepat.

## 2.4 Algoritma Naive Bayes

Naive Bayes merupakan algoritma klasifikasi berbasis probabilistik yang menggunakan Teorema Bayes dengan asumsi bahwa antar fitur bersifat independen. Meskipun asumsi ini tidak selalu terpenuhi dalam dunia nyata, algoritma ini tetap dapat memberikan hasil klasifikasi yang cukup baik.

Teorema Bayes dirumuskan sebagai:  $P(H|X) = P(X|H) * P(H) / P(X)$

Keterangan:

$P(H|X)$ : Probabilitas hipotesis H (misalnya kelas Segar atau Busuk) berdasarkan fitur X.

$P(X|H)$ : Probabilitas fitur X muncul jika diketahui hipotesis H benar.

$P(H)$ : Probabilitas awal dari hipotesis H.

$P(X)$ : Probabilitas awal dari fitur X.

Dalam implementasinya, nilai probabilitas dihitung untuk setiap kelas berdasarkan data latih. Kemudian, kelas dengan probabilitas posterior tertinggi dipilih sebagai hasil klasifikasi.

### 2.4.1 Tahap Training Model Naïve Bayes

#### 1. Perhitungan Probabilitas Awal (Prior)

Untuk setiap kelas K (Segar dan Busuk), dihitung probabilitas Prior:

$$P(K) = \frac{\text{Jumlah data pada kelas } K}{\text{Total jumlah data}}$$

Misal : 55 gambar apel segar dan 55 apel busuk

$$P(K) = \frac{55}{110} = 0.5$$

## 2. Perhitungan Parameter Distribusi Gaussian per Fitur

Naive Bayes di matlab (fitcnb) secara default menggunakan Gaussian

Naive Bayes untuk fitur numerik. Untuk setiap fitur  $x_i$  di setiap kelas  $K$ :

Hitung rata-rata ( $\mu$ )

$$\mu_{i,K} = \frac{\sum_{j=1}^{N_K} x_{i,j}}{N_K}$$

Dimana  $N_K$  = jumlah data pada kelas  $K$ .

Hitung Variasi ( $\sigma^2$ )

$$\sigma_{i,K}^2 = \frac{\sum_{j=1}^{N_K} (x_{i,j} - \mu_{i,K})^2}{N_K}$$

Bentuk distribusi Gaussian :

$$P(x_i | K) = \frac{1}{\sqrt{2\pi\sigma_{i,K}^2}} \cdot \exp\left(-\frac{(x_i - \mu_{i,K})^2}{2\sigma_{i,K}^2}\right)$$

Artinya, setiap fitur diasumsikan mengikuti distribusi normal dalam setiap kelas.

## 3. Asumsi Independensi Fitur

Naive Bayes mengasumsikan bahwa semua fitur independen satu sama lain. Sehingga:

$$P(X|K) = \prod_{i=1}^n P(x_i|K)$$

di mana  $n$  adalah jumlah fitur (dalam sistem ini  $n = 11$ ).

#### 4. Penyimpanan Model

Semua parameter  $P(K) \mu_i, K, \sigma_{i,K}^2$  disimpan dalam objek model.

Model disimpan ke file `model_nb.mat` untuk digunakan saat prediksi.

#### 2.4.2 Tahap Prediksi

##### 1. Ekstraksi fitur dari data uji

Citra diubah menjadi vektor fitur  $X = [x_1, x_2, \dots, x_n]$

##### 2. Perhitungan likelihood

Ambil parameter  $\mu_{i,K}$  dan  $\sigma_{i,K}^2$  dari model.

Hitung peluang kemunculan nilai fitur  $x_i$  dikelas K dengan rumus Gaussian

:

$$P(x_i | K) = \frac{1}{\sqrt{2\pi\sigma_{i,K}^2}} \cdot e^{-\frac{(x_i - \mu_{i,K})^2}{2\sigma_{i,K}^2}}$$

Lakukan untuk semua fitur  $i = 1, 2, \dots, n$

##### 3. Perhitungan Probabilitas Posterior

Hitung

$$P(K | X) \propto P(K) \cdot \prod_{i=1}^n P(x_i | K)$$

$P(K)$  prior probability dari tahap training.

$P(x_i | K)$  nilai Gaussian dari langkah sebelumnya

##### 4. Pemilihan Kelas

Bandingkan nilai  $P(\text{Segar}|X)$  dan  $P(\text{Busuk}|X)$

Pilih kelas dengan probabilitas tertinggi  $\rightarrow$  itulah hasil prediksi.

## **2.5 Validasi Silang (Cross-Validation)**

Cross-validation adalah cara menguji seberapa baik model klasifikasi bisa bekerja pada data yang belum pernah dilihat sebelumnya. Teknik ini membagi dataset menjadi beberapa bagian kecil (disebut fold), kemudian proses belajar dan menguji model dilakukan secara bergantian. Dalam penelitian ini, metode yang digunakan adalah 5-fold cross-validation, dengan langkah-langkah sebagai berikut:

1. Dataset dibagi menjadi lima bagian dengan ukuran yang hampir sama.
2. Pada setiap tahap, empat bagian digunakan untuk melatih model, sedangkan satu bagian digunakan untuk menguji hasilnya.
3. Proses ini diulang lima kali, sehingga setiap bagian data akan menjadi data uji sekali.
4. Akurasi dari masing-masing lima kali pengujian kemudian dijumlahkan dan dibagi rata untuk mendapatkan akurasi akhir.

Manfaat menggunakan cross-validation antara lain:

1. Mengurangi risiko overfitting, yaitu ketika model terlalu cocok hanya dengan data latih.
2. Mengevaluasi sejauh mana model bisa beradaptasi dengan data baru.
3. Memberi gambaran akurasi yang lebih akurat dan stabil.

## **2.6 Matlab dan GUI (Graphical User Interface)**

Matlab adalah perangkat lunak yang digunakan untuk menghitung dan memproses data secara numerik. Banyak orang di bidang teknik dan ilmu

pengetahuan menggunakan Matlab, termasuk dalam memproses gambar digital. Matlab memiliki berbagai fungsi yang sudah siap digunakan untuk memproses gambar, melakukan analisis statistik, dan mengklasifikasikan data. Selain itu, Matlab juga memiliki antarmuka grafis (GUI) yang memudahkan pengguna dalam membuat tampilan interaktif.

Dalam penelitian ini, MATLAB digunakan untuk:

1. Membaca dan menampilkan gambar apel.
2. Melakukan pembersihan data dan mengekstrak fitur secara otomatis.
3. Menggunakan algoritma klasifikasi berbasis Naive Bayes.
4. Memberikan antarmuka interaktif yang mudah digunakan oleh pengguna.

Beberapa fitur GUI yang tersedia adalah:

1. Tombol Training Model: Untuk melatih model berdasarkan data yang sudah disediakan.
2. Tombol Browse Gambar: Untuk memilih gambar secara manual serta mengklasifikasikannya.
3. Tombol Uji Semua: Untuk menguji semua gambar dalam folder uji secara bersamaan.
4. Tombol Crossval Akurasi: Untuk menjalankan proses cross-validation serta menampilkan tingkat akurasi.
5. Tampilan Citra dan Grayscale: Untuk melihat gambar asli dan versi grayscale-nya.
6. Tabel Hasil: Menampilkan hasil klasifikasi, fitur yang terbaca, serta tingkat akurasi (benar/salah).

Dengan adanya antarmuka ini, pengguna seperti petani, pedagang, atau pengelola gudang apel dapat mengoperasikan sistem klasifikasi secara mudah tanpa perlu memahami kode pemrograman. Sistem ini dapat membantu meningkatkan efisiensi proses pengurutan apel berdasarkan kualitas secara otomatis dan objektif.