

# Fruit Image Classification Using Naïve Bayes Algorithm with Histogram of Oriented Gradients (HOG) Feature Extraction

Andika Jodhi Saputra<sup>1</sup>, Widyastuti Andriyani<sup>2\*</sup>

<sup>1,2</sup>Magister Teknologi Informasi, Universitas Teknologi Digital Indonesia, Indonesia, Jl. Raya Janti Jl. Majapahit No.143, Jaranan, Banguntapan, Kec. Banguntapan, Kabupaten Bantul, Daerah Istimewa Yogyakarta 55198

## Informasi Artikel

Diterima : 25 Februari 2025  
Revisi : 28 Februari 2025  
Publikasi : 20 Maret 2025

## Kata Kunci:

Naïve Bayes  
Klasifikasi Buah  
Histogram of Oriented Gradients (HOG)  
Pemrosesan Gambar

## ABSTRAK

Sistem klasifikasi menggunakan algoritma Naïve Bayes dikembangkan untuk membedakan antara buah segar dan busuk, khususnya apel, pisang, dan jeruk. Penelitian ini memanfaatkan dataset yang terdiri dari 13.599 gambar dan menerapkan teknik Histogram of Oriented Gradients (HOG) untuk ekstraksi fitur, diikuti dengan pelatihan dan evaluasi model. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mencapai akurasi sebesar 87%, dengan presisi tertinggi pada kelas apel segar (0,9792) dan recall tertinggi pada kelas apel busuk (0,9843). Kelas pisang busuk menunjukkan kinerja seimbang dengan F1-score tertinggi sebesar 0,9085. Meskipun terdapat beberapa kesalahan klasifikasi, terutama pada kategori buah jeruk busuk, penelitian ini menunjukkan bahwa teknik pengolahan citra memiliki potensi besar dan dapat diandalkan untuk menilai kualitas buah berdasarkan karakteristik visual.

## ABSTRACT

A classification system using Naïve Bayes algorithm was developed to distinguish between fresh and rotten fruits, specifically apples, bananas and oranges. This research utilized a dataset consisting of 13,599 images and applied the Histogram of Oriented Gradients (HOG) technique for feature extraction, followed by model training and evaluation. The results showed that the Naïve Bayes algorithm achieved an accuracy of 87%, with the highest precision in the fresh apple class (0.9792) and the highest recall in the rotten apple class (0.9843). The rotten banana class showed a balanced performance with the highest F1-score of 0.9085. Although there were some misclassifications, especially in the rotten citrus fruit category, this study shows that image processing techniques have great potential and are reliable for assessing fruit quality based on visual characteristics.

This is an open-access article under the [CC BY-SA](#) license



## \*Penulis Koresponden

Email: [student.andikajodhi@mti.utdi.ac.id](mailto:student.andikajodhi@mti.utdi.ac.id)

Cara sitasi IEEE::

A. J. Saputra dan W. Andriyani, "Fruit image classification using Naïve Bayes algorithm with histogram of oriented gradients (HOG) feature extraction," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 1, pp. 215-227, Maret 2025. doi: 10.30811/jaise.v5i1.6536

## 1. PENDAHULUAN

Buah-buahan merupakan salah satu kategori komoditas pertanian terpenting di Indonesia. Permintaan akan buah-buahan cukup tinggi, terlihat dari berbagai macam buah-buahan yang tersedia di pasar modern dan tradisional. Apel, pisang, dan jeruk merupakan beberapa buah yang paling digemari oleh masyarakat Indonesia, mulai dari yang muda hingga yang tua. Hal ini menunjukkan bahwa ketiga jenis buah tersebut telah banyak dikonsumsi oleh masyarakat dan memiliki daya saing yang baik. Tingginya produksi dan distribusi buah-

buah tersebut mengharuskan para petani untuk dapat mengidentifikasi kesegaran buah-buahan tersebut dan mengurangi risiko pembusukan [1].

Sejalan dengan kemajuan zaman, teknologi saat ini berkembang dengan pesat. Penggunaan pengolahan citra untuk mengidentifikasi gambar semakin hari semakin meningkat di berbagai bidang seperti pengolahan citra industri, pencitraan medis, pencitraan real time, klasifikasi tekstur, pengenalan objek, dan sebagainya [2]. Pengolahan citra dan visi komputer di bidang pertanian merupakan bidang yang sangat luas dan memiliki banyak aplikasi dalam berbagai aspek kehidupan sehari-hari. Bidang ini mencakup berbagai macam operasi, termasuk peningkatan, klasifikasi, pengenalan, dan deteksi pada gambar dan video yang dihasilkan dari kegiatan sehari-hari [3]. Machine Learning, yang merupakan cabang dari Artificial Intelligence (AI), telah membawa revolusi besar dalam berbagai bidang teknologi [4] Metode ini diklasifikasikan menjadi tiga jenis utama berdasarkan cara kerjanya: supervised, unsupervised, dan semi-supervised [5]. Di antara metode klasifikasi tersebut, Naïve Bayes, dengan memanfaatkan kekuatan AI, telah menjadi favorit [6]. Kesederhanaan dan keakuratannya dalam memprediksi kategori data membuatnya menjadi pilihan utama dalam aplikasi Machine Learning. Metode ini paling cocok untuk kumpulan data yang lebih besar yang mungkin berisi jutaan gambar atau sampel data. Keunggulan: Metode ini melibatkan kriteria prediksi yang sederhana, cepat, dan mudah [7].

Kualitas buah memiliki dampak yang signifikan terhadap daya tarik konsumen, terutama dalam kaitannya dengan bentuk buah itu sendiri. Biasanya, kualitas buah dinilai dengan pengamatan visual. Namun, metode ini kurang efisien dan memakan waktu. Salah satu solusi untuk masalah ini adalah penggunaan analisis citra digital, yang melibatkan teknik pemrosesan dan analisis citra berbantuan computer [8]. Sebagai contoh, klasifikasi buah apel, jeruk, dan pisang dapat dilakukan dengan mempertimbangkan fitur warna dan nilai RGB untuk membedakan jenis dan kondisinya (segar atau busuk). Sebagai contoh, warna buah apel dipengaruhi oleh pigmen pada kulitnya. RGB (Red-Green-Blue) adalah model warna yang menghasilkan warna yang berbeda dengan mengkombinasikan warna merah, hijau, dan biru dengan proporsi yang berbeda.

Permasalahan tersebut dapat diselesaikan dengan menggunakan beberapa metode dari bidang matematika, diantaranya adalah metode Naïve Bayes dan HOG (Histogram of Oriented Gradients) [9]. Histogram of Oriented Gradients (HOG) merupakan sebuah metode yang digunakan pada image processing bertujuan untuk deteksi obyek. Teknik ini menghitung nilai gradien dalam daerah tertentu suatu citra. Tiap citra mempunyai karakteristik yang di tunjukkan oleh distribusi gradien [10]. Metode HOG robust terhadap variasi pencahayaan dan mampu mencapai tingkat akurasi yang tinggi dalam mendeteksi berbagai objek dengan tekstur [11]. HOG merupakan deskriptor fitur berbasis gradien yang efisien untuk identifikasi data dan menunjukkan kinerja yang sangat baik dibandingkan dengan metode ekstraksi fitur lainnya [9], Proses ekstraksi fitur HOG dimulai dengan mengubah gambar berwarna menjadi skala abu-abu, yang membantu mengurangi informasi warna. Selanjutnya, nilai gradien pencahayaan dihitung untuk setiap piksel, diikuti dengan histogram orientasi gradien untuk setiap blok. Proses ini memungkinkan pengambilan fitur yang kuat dengan adanya perubahan bentuk, dan nilai fitur yang dinormalisasi kemudian diperoleh untuk setiap blok.

## 2. METODE

Pada penelitian ini dilakukan melalui beberapa langkah, dimulai dengan pengumpulan data citra buah dari dataset "Fruits Fresh and Rotten for Classification" di Kaggle, yang terdiri dari 13.599 gambar. Setelah itu, citra diolah dengan mengubahnya menjadi grayscale dan mengekstraksi fitur menggunakan Histogram of Oriented Gradients (HOG). Data kemudian dibagi menjadi dua bagian: 80% untuk data latih dan 20% untuk data uji, di mana data latih digunakan untuk melatih model Naïve Bayes dalam mengklasifikasikan buah segar dan busuk. Evaluasi model dilakukan dengan menghitung akurasi, precision, dan recall berdasarkan confusion matrix untuk memastikan kemampuan generalisasi model terhadap data baru.

### 2.1. Pengumpulan Data

Data yang disajikan dalam penelitian ini berasal dari dataset "Fruits Fresh and Rotten for Classification", yang tersedia di situs web Kaggle (<https://www.kaggle.com/datasets/sriramr/fruits-fresh-and-rotten-for-classification>). Dataset ini terdiri dari total 13.599 gambar buah-buahan segar dan busuk, yang disajikan dalam format gambar .png. Spesifikasi dataset yang digunakan ditunjukkan pada Tabel 1. Data train digunakan untuk melatih model naive bayes, sedangkan data test digunakan untuk mengevaluasi kinerja model naive bayes. Pembagian data ini penting untuk memastikan bahwa model dapat menggeneralisasi dengan baik terhadap data baru. Dalam machine learning, dataset dibagi menjadi dua bagian: 80% untuk data train dan 20% untuk data test. Pembagian data dapat dilihat dibawah ini:

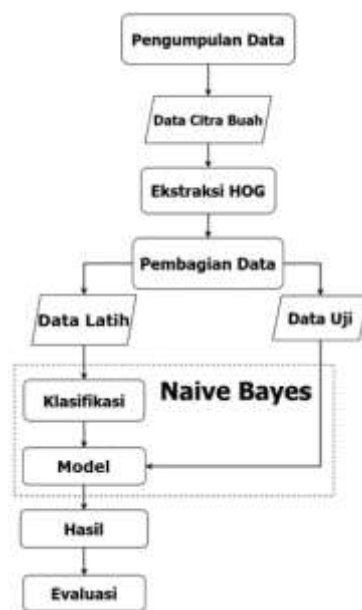
1. Total jumlah data train = 10.901
2. Total jumlah data test = 2.698
3. Total dataset = 13.599

Tabel 1. Dataset Citra Buah

No	Nama Kelas	Data Train	Data Test
1	freshapples	1,693	395
2	freshbanana	1,581	381
3	freshoranges	1,466	388
4	rottenapples	2,342	601
5	rottenbanana	224	530
6	rottenoranges	1,595	403
	Jumlah	10,901	2,698
	Total Keseluruhan	13,599	

## 2.2. Model yang Diusulkan

Model yang diusulkan untuk prediksi citra buah busuk dan buah segar dalam penelitian ini ditunjukkan pada Gambar 1. Yang pertama pengumpulan data, pra-pemrosesan, pembangunan model, pengujian model, dan evaluasi hasil. Proses dimulai dengan pengumpulan data citra buah, diikuti oleh ekstraksi grayscale menggunakan fitur Histogram of Oriented Gradients (HOG) [12]. Setelah ekstraksi, data dibagi menjadi dua bagian: data train dan data test. Data train digunakan untuk klasifikasi antara buah segar dan buah busuk, yang berfungsi sebagai pembanding untuk data test. Selanjutnya, data test dianalisis dengan membandingkannya menggunakan model Naive Bayes.



Gambar 1. Model yang Diusulkan

Model Naïve Bayes adalah salah satu metode yang digunakan untuk mengklasifikasikan data. Klasifikasi Bayesian merupakan pendekatan statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas (class) [11]. Teorema ini dikemukakan oleh ilmuwan Inggris, Thomas Bayes, yang mengembangkan metode untuk memprediksi peluang di masa depan berdasarkan pengalaman dan data yang diperoleh di masa lalu [13]. Teorema Bayes ini kemudian dikombinasikan dengan istilah "Naive," yang merujuk pada asumsi bahwa setiap atribut atau variabel bersifat independen satu sama lain [14]. Dalam konteks frekuentis, model ini mengartikan representasi probabilitas sebagai invers dari probabilitas melalui dua skenario kejadian. Naïve Bayes merupakan salah satu metode dalam machine learning yang menggunakan perhitungan probabilitas. Algoritma ini memanfaatkan prinsip probabilitas dan statistik sederhana dengan asumsi bahwa antar atribut dalam kelas tidak saling tergantung (independen). Dasar dari Naïve Bayes yang digunakan dalam pemrograman adalah persamaan Bayes yang dinyatakan pada persamaan 1.

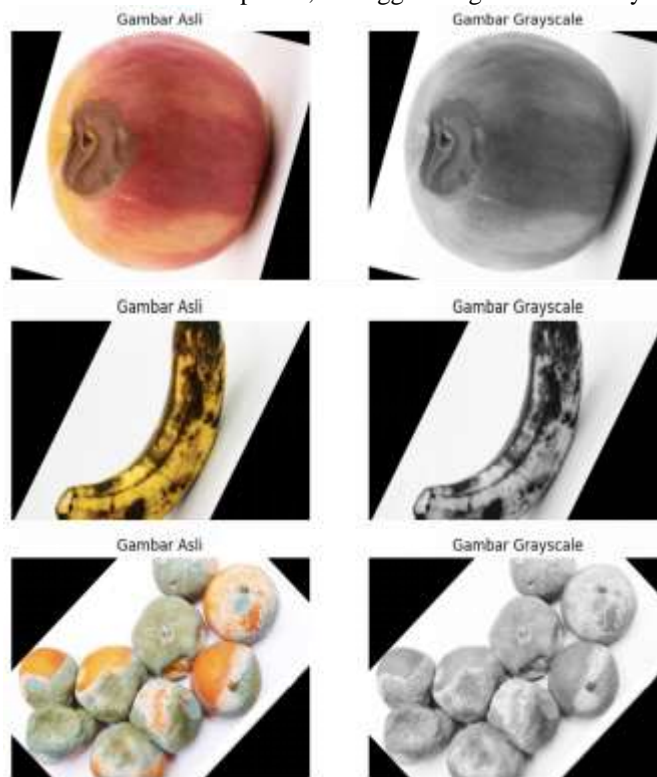
$$P(h|d) = \frac{P(d|h).P(h)}{P(d)} \quad (1)$$

Dimana:

- $d$  = Data kelas yang belum diketahui
- $h$  = Hipotesis data suatu kelas
- $(P(h|d))$  = Probabilitas hipotesis ( $h$ ) jika diberikan data ( $d$ ).
- $(P(d|h))$  = Probabilitas data ( $d$ ) jika hipotesis ( $h$ ) benar.
- $(P(h))$  = Probabilitas hipotesis ( $h$ ) sebelum melihat data.
- $(P(d|P(d)))$  = Probabilitas data ( $d$ )

### 2.3. Model yang Diusulkan

Histogram of Oriented Gradients (HOG) adalah metode ekstraksi fitur yang efektif dalam pemrosesan gambar, terutama untuk aplikasi visi komputer dan pengenalan objek. Metode ini tahan terhadap perubahan pencahayaan dan mampu mencapai akurasi yang tinggi dalam mendeteksi objek bertekstur [11]-[15]. Proses ekstraksi HOG dimulai dengan mengubah citra berwarna menjadi grayscale, yang bertujuan untuk meminimalisir informasi warna. Selanjutnya, nilai gradien pencahayaan dari setiap piksel dihitung, dan histogram orientasi gradien dibuat untuk setiap blok, sehingga menghasilkan fitur yang kuat [2].



Gambar 2 Citra Buah Grayscale

Mengekstrak fitur HOG dari dataset setiap kelas pada citra buah. Prosesnya dimulai dengan mengubah gambar berwarna menjadi skala abu-abu untuk mengurangi kompleksitas data. Setelah itu, fitur HOG dihitung dengan menggunakan parameter yang ditentukan, yaitu ukuran piksel per sel dan jumlah sel per blok, rumusnya dapat dilihat dibawah ini:

$$\text{Magnitude} = \sqrt{G_x^2 + G_y^2} \quad (2)$$

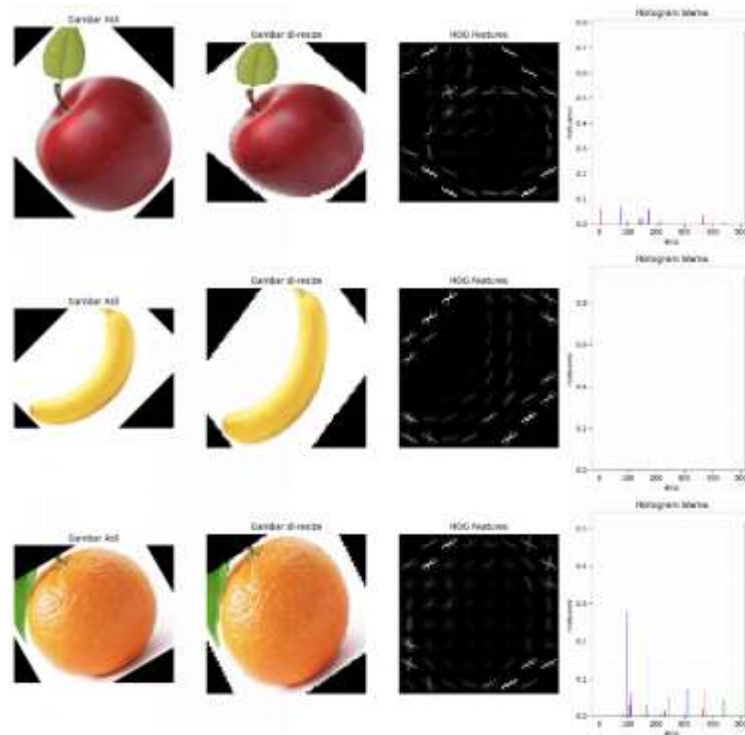
$$\text{Orientation} = \arctan \frac{G_y}{G_x} \quad (3)$$

Hasil dari fungsi ini adalah vektor fitur yang merepresentasikan informasi orientasi gradien dari gambar.

Menghitung histogram warna dari gambar. Histogram dihitung dengan mempertimbangkan tiga saluran warna (merah, hijau, dan biru) dan dibagi menjadi 8 bin untuk masing-masing saluran, berikut ini adalah rumus untuk penggabungan histogram dari 3 saluran dan normalisasi:

$$\text{Nilai Normalisasi} = \frac{\text{Nilai Bin}}{\text{Total Piksel}} \quad (4)$$

Setelah histogram dihitung, proses normalisasi dilakukan untuk memastikan bahwa nilai histogram berada dalam rentang yang konsisten, dan hasilnya diratakan menjadi satu dimensi untuk memudahkan analisis lebih lanjut.



Gambar 3 Histogram Warna Citra Buah

#### 2.4. Klasifikasi

Implementasi model yang telah dilatih untuk memproses data baru. Data yang belum dikenali di ekstraksi fitur HOG dan diinputkan ke dalam model Naive Bayes. Model memprediksi label kelas berdasarkan probabilitas posterior tertinggi. Hasil klasifikasi digunakan dalam berbagai aplikasi, seperti pengenalan objek, pengelompokan gambar, atau sistem deteksi otomatis [16]. Pada penelitian ini, hasil klasifikasi divisualisasikan untuk memudahkan interpretasi, serta dibandingkan dengan label ground truth untuk validasi akhir.

Dalam penelitian ini, evaluasi kinerja model dilakukan dengan menggunakan metrik utama seperti akurasi, precision, dan recall. Metrik-metrik ini dihitung berdasarkan analisis confusion matrix, yang memberikan gambaran tentang seberapa baik model dapat mengklasifikasikan data dengan benar. Akurasi mengukur persentase prediksi yang benar dari keseluruhan data, precision menilai ketepatan prediksi positif, dan recall mengukur kemampuan model dalam mendeteksi semua kasus positif. Rumus-rumus yang digunakan untuk menghitung metrik ini adalah sebagai berikut :

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \times 100\% \quad (5)$$

$$Presisi = \frac{TP}{TP+FP} \quad (6)$$

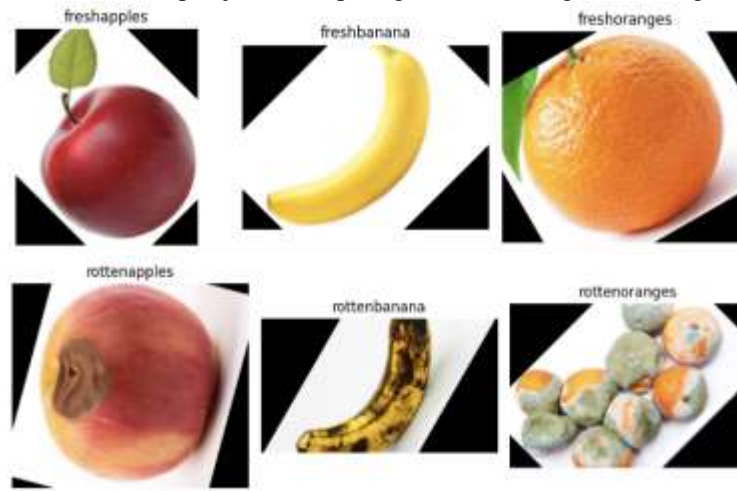
$$Recall = \frac{TP}{TP+FN} \quad (7)$$

Keterangan:

- TP (True Positive) adalah jumlah prediksi positif yang benar
- TN (True Negative) adalah jumlah prediksi negatif yang benar,
- FP (False Positive) adalah jumlah prediksi positif yang salah,
- FN (False Negative) adalah jumlah prediksi negatif yang salah.

### 3. HASIL DAN PEMBAHASAN

Untuk mengimplementasikan model yang diusulkan, bahasa pemrograman Python dipilih sebagai alat untuk proses klasifikasi citra buah apel, jeruk, dan pisang ke dalam kategori buah segar dan busuk.



Gambar 4 Citra Buah Segar dan Buah Busuk

#### 3.1. Ekstraksi Fitur HOG (Histogram of Oriented Gradients)

Histogram of Oriented Gradients (HOG) adalah metode yang umum digunakan dalam pengolahan citra dan pengenalan pola, khususnya untuk deteksi objek[10]. Metode ini dimulai dengan mengonversi citra ke dalam skala abu-abu, diikuti dengan perhitungan gradien menggunakan operator Sobel untuk mengidentifikasi arah dan kekuatan perubahan intensitas. Selanjutnya, histogram orientasi dibentuk dengan membagi citra menjadi sel-sel kecil dan menghitung distribusi arah gradien dalam setiap sel, yang kemudian dinormalisasi untuk meningkatkan ketahanan terhadap variasi pencahayaan. Hasil ekstraksi fitur HOG ini memberikan representasi yang kuat dari objek dalam dataset, yang dapat digunakan untuk pelatihan model deteksi objek. Proses ekstraksi HOG dilakukan dengan langkah-langkah sebagai berikut:

##### 1. Pra-pemrosesan Citra

Sebelum melakukan ekstraksi fitur Histogram of Oriented Gradients (HOG), citra harus melalui proses pra-pemrosesan yang meliputi konversi ke format grayscale untuk mengurangi kompleksitas data, normalisasi ukuran untuk memastikan konsistensi dimensi, dan penghapusan noise yang tidak diinginkan menggunakan teknik seperti filter Gaussian atau median.



Gambar 5 Konversi Citra ke Format Grayscale

##### 2. Perhitungan Gradien

Untuk setiap sel, gradien citra dihitung menggunakan operator Sobel

$$G_x = I(x + 1, y) - I(x - 1, y), \quad G_y = I(x, y + 1) - I(x, y - 1) \quad (8)$$

Gradien ini memberikan informasi tentang arah dan kekuatan perubahan intensitas piksel.

##### 3. Penghitungan Histogram

Citra dibagi menjadi sel berukuran 8x8 piksel:

$$\begin{bmatrix} 181 & 178 & 175 & 176 & 176 & 179 & 180 & 179 \\ 184 & 180 & 176 & 176 & 177 & 181 & 183 & 179 \\ 185 & 182 & 178 & 178 & 178 & 181 & 182 & 179 \\ 185 & 183 & 182 & 180 & 178 & 179 & 179 & 178 \\ 184 & 183 & 183 & 180 & 179 & 178 & 177 & 178 \\ 183 & 182 & 182 & 180 & 178 & 178 & 178 & 178 \\ 181 & 181 & 181 & 179 & 178 & 178 & 178 & 178 \\ 179 & 179 & 179 & 178 & 178 & 178 & 178 & 178 \end{bmatrix}$$

Hasil perhitungan gradien:

$$G_x = -29$$

$$G_y = 15$$

$$G = \sqrt{(-29)^2 + (15)^2} = \sqrt{841 + 225} = \sqrt{1066} = 32,68$$

Setiap sel berukuran 8x8 piksel menghasilkan histogram arah gradien dengan membagi arah gradien ke dalam 9 bin dengan rentang hingga bertujuan untuk menangkap informasi lokal dari citra

Arah gradien :

$$\theta = \text{atan2}(15, -29) = 2,74$$

Konversi ke derajat:

$$\Theta = 2,74 \times 180 = 157,3$$

Hasil histogram yaitu berada dalam bin 8 () dan magnitudo gradien 32,68 di input ke dalam bin 8.

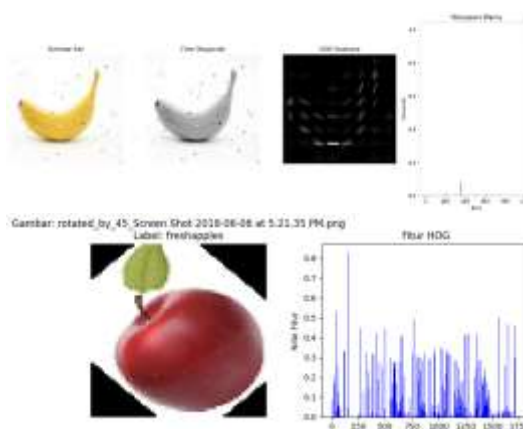
#### 4. Normalisasi Blok

Blok-blok sel yang berdekatan (misalnya 2 x 2 sel) dinormalisasi untuk meningkatkan ketahanan terhadap perubahan pencahayaan dan kontras. Hasil histogram normalisasi untuk Blok (0, 0) dan (0, 2) adalah:

Histogram Normalisasi = [0,7247,0,0,0,1563,0,2635,0,3038,0,5378,0]

#### 5. Penggabungan Fitur

Semua blok fitur digabungkan menjadi satu vektor akhir untuk seluruh gambar.



Gambar 6 Hasil Ekstrak Fitur HOG

Pada Gambar 6, ditampilkan sebuah apel yang telah diputar sejauh 45 derajat dan diubah ukurannya menjadi 64 x 64 piksel. Proses ini bertujuan untuk memastikan konsistensi dalam analisis citra. Dari gambar tersebut, fitur Histogram of Oriented Gradients (HOG) dan histogram warna diekstraksi untuk membentuk dataset yang representatif. Fitur HOG memberikan informasi tentang orientasi dan bentuk objek, sementara histogram warna menangkap distribusi warna dalam citra. Dataset yang dihasilkan dari proses ekstraksi ini kemudian digunakan untuk melatih model klasifikasi yang bertujuan membedakan antara buah segar dan buah busuk. Dengan demikian, pendekatan ini memungkinkan pengembangan sistem yang lebih akurat dalam pengenalan dan klasifikasi kualitas buah.

### 3.2. Training dan Prediksi dengan Model Gaussian Naïve Bayes

Gaussian Naive Bayes adalah metode klasifikasi probabilistik yang mengasumsikan independensi antar fitur dan menggunakan distribusi Gaussian untuk memodelkan data kontinu. Prosesnya melibatkan perhitungan mean dan varians untuk setiap fitur pada masing-masing kelas, serta menghitung probabilitas posterior untuk menentukan prediksi kelas dengan probabilitas tertinggi.

#### 1. Rumus Distribusi Normal

Probabilitas untuk setiap fitur pada kelas Ck dihitung menggunakan rumus distribusi normal:

$$P(X_i|C_k) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(X_i - \mu_i)^2}{2\sigma_i^2}\right) \tag{9}$$

Keterangan:

- $X_i$  = nilai fitur yang diujikan.
- $\mu_i$  = rata-rata fitur i untuk kelas Ck
- $\sigma_i^2$  = varians fitur i untuk kelas Ck

#### 2. Data Mean dan Varians untuk Masing-Masing Kelas

Dalam analisis klasifikasi buah, data yang disajikan mencakup rata-rata ( $\mu$ ) dan varians ( $\sigma^2$ ) dari dua fitur yang relevan untuk enam kelas buah yang berbeda, yaitu *freshapples*, *freshbanana*, *freshoranges*, *rottenapples*, *rottenbanana*, dan *rottenoranges*. Rata-rata fitur pertama dan kedua memberikan gambaran tentang karakteristik fisik dari masing-masing jenis buah, sementara varians menunjukkan seberapa besar variasi dalam data untuk setiap fitur

Tabel 2 Data Mean dan Varians

Kelas	$\mu_1$ (fitur 1)	$\mu_2$ (fitur 2)	$\sigma_1^2$ (varians fitur 1)	$\sigma_2^2$ (varians fitur 2)
freshapples	0.07155103	0.02219062	0.01816059	0.00560215
freshbanana	0.09357712	0.01829614	0.02791861	0.00383337
freshoranges	0.07967494	0.02132208	0.02220574	0.00497054
rottenapples	0.07094855	0.02650812	0.01577268	0.00442407
rottenbanana	0.08729977	0.01957312	0.02282626	0.00386834
rottenoranges	0.07822917	0.02116891	0.01742324	0.00350419

Kelas *freshapples* :

$$P(X|C_{freshapples}) = P(X_1|C_{freshapples}) \times P(X_2|C_{freshapples}) = 0,288 \times 0,539 = 0,153$$

Kelas *freshbanana* :

$$P(X|C_{freshbanana}) = P(X_1|C_{freshbanana}) \times P(X_2|C_{freshbanana}) = 0,299 \times 0,645 = 0,147$$

Kelas *freshoranges*:

$$P(X|C_{freshoranges}) = P(X_1|C_{freshoranges}) \times P(X_2|C_{freshoranges}) = 0,261 \times 0,566 = 0,148$$

Kelas *rottenapples* :

$$P(X|C_{rottenapples}) = P(X_1|C_{rottenapples}) \times P(X_2|C_{rottenapples}) = 0,305 \times 0,592 = 0,180$$

Kelas *rottenbanana*

$$P(X|C_{rottenbanana}) = P(X_1|C_{rottenbanana}) \times P(X_2|C_{rottenbanana}) = 0,256 \times 0,654 = 0,165$$

Kelas *rottenoranges*

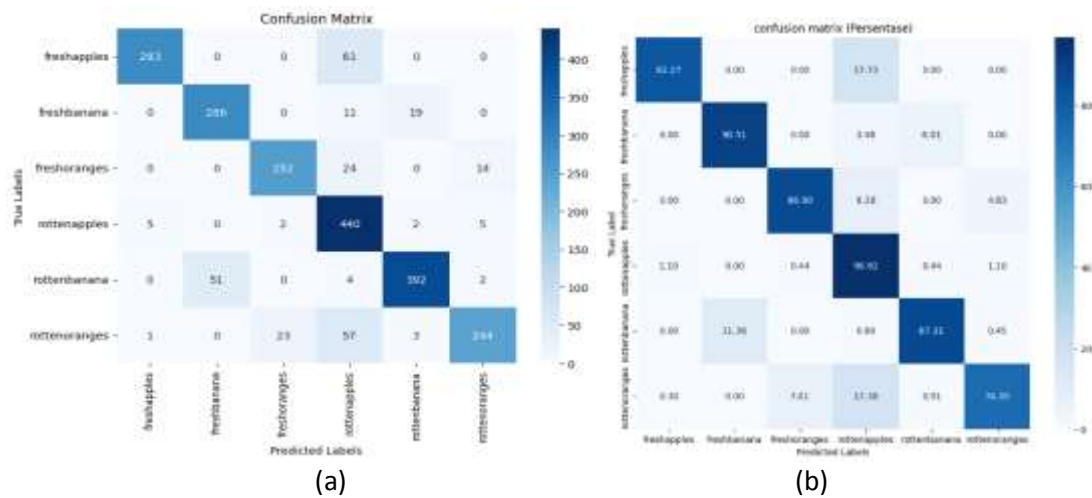
$$P(X|C_{rottenoranges}) = P(X_1|C_{rottenoranges}) \times P(X_2|C_{rottenoranges}) = 0,293 \times 0,670 = 0,197$$

Tabel 3 Hasil Probabilitas

No	Nama Kelas	Probabilitas Posterio
1	freshapples	0,153
2	freshbanana	0,147
3	freshoranges	0,148
4	rottenapples	0,180
5	rottenbanana	0,165
6	rottenoranges	0,197

Hasil perhitungan probabilitas posterior dalam model Naïve Bayes menunjukkan bahwa kelas rottenoranges memiliki probabilitas tertinggi sebesar 0,197, diikuti oleh rottenapples dengan 0,180 dan rottenbanana 0,165, mengindikasikan bahwa buah-buah ini lebih mungkin tergolong sebagai buah busuk. Sementara itu, kelas freshapples (0,153), freshoranges (0,148), dan freshbanana (0,147) menunjukkan probabilitas lebih rendah, menandakan bahwa fitur yang diukur lebih mendukung klasifikasi buah sebagai busuk daripada segar.

### 3.3. Evaluasi Model Naïve Bayes



Gambar 7 Confusion Matrix Citra Buah untuk (a) Jumlah citra dan (b) Persentase

Klasifikasi data adalah proses yang krusial dalam evaluasi model klasifikasi, di mana kinerja model diukur menggunakan berbagai metrik, termasuk akurasi, precision, recall, F1-score, dan confusion matrix. Tujuan dari evaluasi ini adalah untuk menilai kemampuan model Gaussian Naive Bayes dalam mengklasifikasikan data buah ke dalam enam kategori: freshapples, freshbanana, freshoranges, rottenapples, rottenbanana, dan rottenoranges.

Berikut adalah langkah-langkah perhitungan metrik klasifikasi (precision, recall, F1-score) berdasarkan confusion matrix untuk masing-masing kelas:

#### 1. Kelas freshapples

$$Precision = \frac{283}{283 + 61} = 0,9792$$

$$Recall = \frac{283}{283 + 61} = 0,8225$$

$$F1 - Score = 2 \times \frac{0,9792 \times 0,8225}{0,9792 + 0,8225} = 0,8945$$

#### 2. Kelas freshbanana

$$Precision = \frac{286}{286 + 51} = 0.8487$$

$$Recall = \frac{286}{286 + 30} = 0.9050$$

$$F1 - Score = 2 \times \frac{0.8487 \times 0.9050}{0.8487 + 0.9050} = 0.8759$$

#### 3. Kelas freshoranges

$$Precision = \frac{252}{252 + 25} = 0.9097$$

$$Recall = \frac{252}{252 + 38} = 0.8694$$

$$F1 - Score = 2 \times \frac{0.9097 \times 0.8694}{0.9097 + 0.8694} = 0.8890$$

4. Kelas rottenapples

$$Precision = \frac{440}{440 + 157} = 0.7368$$

$$Recall = \frac{440}{440 + 7} = 0.9843$$

$$F1 - Score = 2 \times \frac{0.7368 \times 0.9843}{0.7368 + 0.9843} = 0.8429$$

5. Kelas rottenbanana

$$Precision = \frac{392}{392 + 24} = 0.9422$$

$$Recall = \frac{392}{392 + 55} = 0.8771$$

$$F1 - Score = 2 \times \frac{0.9422 \times 0.8771}{0.9422 + 0.8771} = 0.9085$$

6. Kelas rottenoranges

$$Precision = \frac{244}{244 + 21} = 0.9207$$

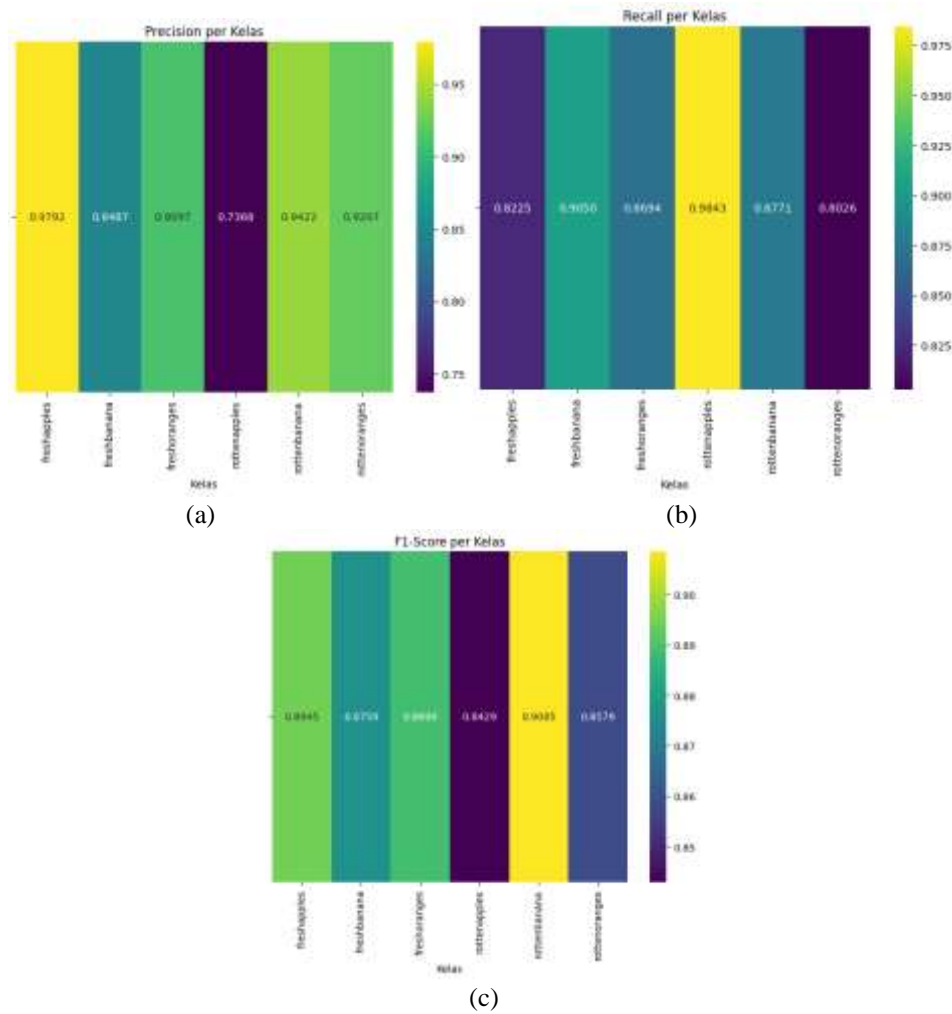
$$Recall = \frac{244}{244 + 60} = 0.8026$$

$$F1 - Score = 2 \times \frac{0.9207 \times 0.8026}{0.9207 + 0.8026} = 0.8579$$

Pada gambar 8 matriks klasifikasi yaitu hasil dari evaluasi model dengan menggunakan tiga matrik utama yaitu Precision, Recall, dan F1- Score yang dihitung pada tiap kelas. Setiap grafik memberikan perhitungan gambaran model melakukan klasifikasi terhadap berbagai kelas pada data citra.

Pada gambar 8 (a) pada grafik menampilkan Precision per kelas, kelas freshapple mendapatkan nilai precision tertinggi sebesar 0,9792 yang menunjukkan model sangat akurat dalam melakukan klasifikasi citra buah apel segar. Sedangkan di kelas rottenapple menghasilkan precision terendah yaitu 0,7368 menandakan bahwa banyak data citra *rottenapple* salah dikategorikan ke kelas lain. Hal ini mengindikasikan adanya kebingungan atau ambiguitas antara kelas-kelas, karena fitur-fitur citra yang kualitas gambarnya buruk atau keterbatasan model dalam membedakan antara kelas yang mirip sehingga, meskipun banyak gambar terdeteksi *rottenapple*, kesalahan dalam klasifikasi yang terjadi menyebabkan nilai precision menjadi rendah, meskipun kelas rottenapple mendapatkan recall yang tinggi.

Pada gambar 8 (b) grafik recall per kelas. Kelas rottenbanana mendapatkan nilai recall tertinggi 0,9843 yang menunjukkan bahwa model mampu menangkap Sebagian besar kelas rottenbanana. Sebaliknya untuk kelas, rottenoranges, menunjukkan recall paling rendah 0,8026 yang mengidentifikasi bahwa beberapa contoh dari kelas rottenorange tidak terdeteksi dengan baik oleh model faktor yang mungkin menyebabkan hal ini adalah adanya kesamaan visual antara gambar rottenorange dengan kelas lain, seperti rottenbanana atau freshorange, yang membuat model kesulitan membedakan kelas-kelas tersebut.



Gambar 8: Matriks klasifikasi untuk (a) Precision per kelas, (b) Recall per kelas, dan (c) F1-Score per kelas.

Pada gambar 8 (c) tampilan grafik F1-Score per kelas, yang menghasilkan gambaran tentang keseimbangan antara precision dan recall. Kelas rottenbanana mendapatkan nilai F1-Score tertinggi (0,9085), yang mengidentifikasi kinerja yang seimbang dan baik antara precision dan recall. Kelas freshapples juga memiliki F1-Score yang tinggi 0,8945, menunjukkan keseimbangan yang baik meskipun precision dan recall-nya tidak setinggi kelas rottenbanana.

Tabel 4 Hasil Klasifikasi

Kelas	Precision	Recall	F1-Score
freshapples	0.9792	0.8225	0.8945
freshbanana	0.8487	0.9050	0.8759
freshoranges	0.9097	0.8694	0.8890
rottenapples	0.7368	0.9843	0.8429
rottenbanana	0.9422	0.8771	0.9085
rottenoranges	0.9207	0.8026	0.8579

Tingkat akurasi adalah proporsi dari prediksi yang benar dibandingkan dengan total prediksi yang dilakukan. Dalam konteks ini, akurasi dapat dihitung dengan rumus berikut:

$$Akurasi = \frac{True\ Positives\ (TP)}{Total\ Prediksi}$$

Jumlah elemen diagonal:

- freshapples = 283
- freshbanana = 286
- freshoranges = 252

- rottenapples = 440
- rottenbanana = 392
- rottenoranges = 244

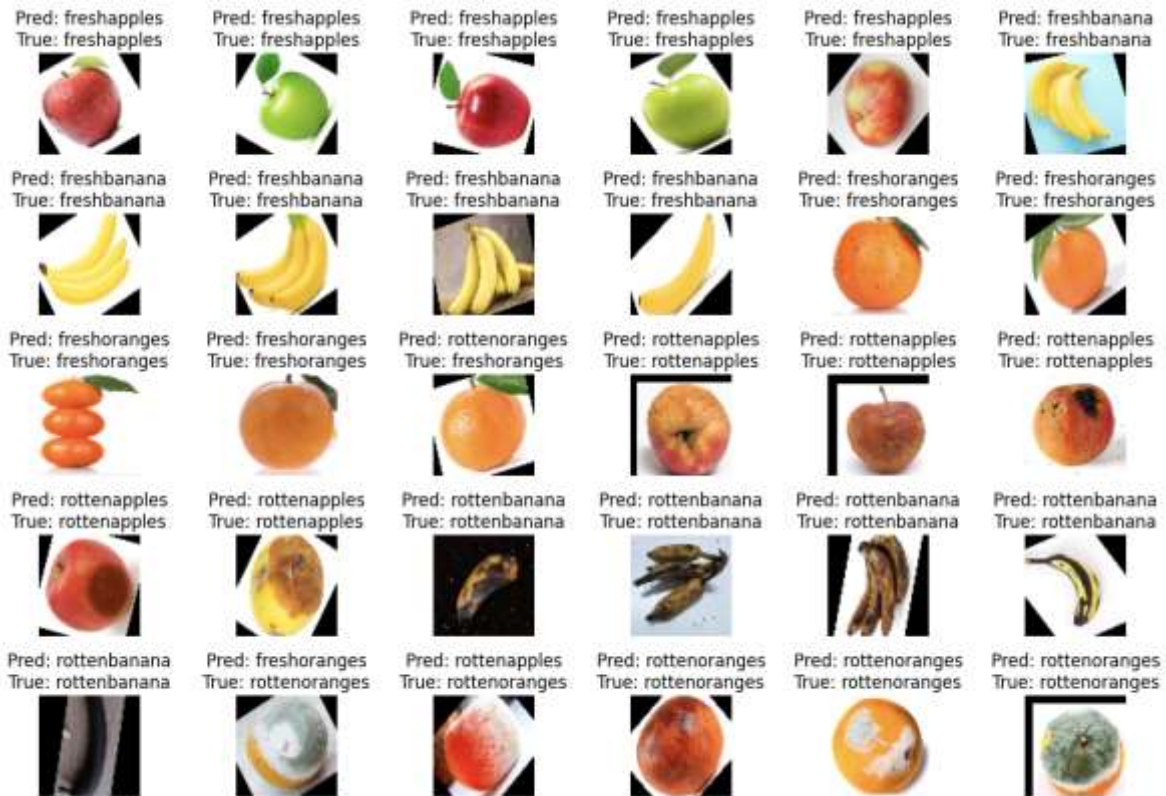
True Positives (TP) = 283 + 286 + 252 + 440 + 392 + 244 = 1897

Total Prediksi (TP + FP + FN):

Total prediksi = 283 + 0 + 0 + 61 + 0 + 0 + 0 + 286 + 0 + 11 + 19 + 0 + 0 + 0 + 252 + 24 + 0 + 14 + 5 + 0 + 2 + 440 + 2 + 5 + 0 + 51 + 0 + 4 + 392 + 2 + 1 + 0 + 23 + 57 + 3 + 244 = 2181

$$\text{Akurasi} \frac{1897}{2181} = 0,87\%$$

Berdasarkan perhitungan akurasi model Naive Bayes dalam klasifikasi buah segar dan buah busuk, diperoleh hasil yang menunjukkan efektivitas model dalam mengidentifikasi kategori buah dengan tepat. Dalam analisis ini, model berhasil mengklasifikasikan sejumlah besar data dengan akurasi mencapai 87%, yang mencerminkan kemampuan model untuk membedakan antara buah segar, seperti *freshapples*, *freshbanana* dan *freshoranges* serta buah busuk, seperti *rottenapples*, *rottenbanana* dan *rottenoranges*. Hasil ini tidak hanya menegaskan keandalan metode Naive Bayes dalam pengolahan data klasifikasi, tetapi juga memberikan wawasan berharga untuk aplikasi lebih lanjut dalam bidang pertanian dan distribusi produk pangan.



Gambar 9 Sampel prediksi citra

#### 4. KESIMPULAN

Algoritma Naïve Bayes efektif dalam klasifikasi citra buah segar dan busuk dengan akurasi 87% dari 13.599 gambar, khususnya pada jenis apel, pisang, dan jeruk. Kelas *freshapples* menunjukkan precision tertinggi (0,9792) dan recall (0,8225), sementara *rottenapples* memiliki recall tertinggi (0,9843), meskipun precision-nya lebih rendah (0,7368). Probabilitas posterior menunjukkan bahwa kelas *rottenoranges* memiliki probabilitas tertinggi sebesar 0,197, diikuti oleh *rottenapples* dan *rottenbanana*. Ekstraksi fitur menggunakan Histogram of Oriented Gradients (HOG) berhasil meningkatkan akurasi klasifikasi dengan menangkap karakteristik tekstur dan bentuk buah. Meskipun demikian, terdapat beberapa kesalahan klasifikasi, terutama pada *rottenoranges* (recall 0,8026). Namun, model ini memiliki keterbatasan, seperti asumsi independensi antar

fitur yang tidak berlaku pada citra buah, di mana fitur seperti tekstur dan warna saling berkorelasi. Selain itu, distribusi fitur yang tidak selalu Gaussian dapat memengaruhi akurasi. Oleh karena itu, disarankan untuk mempertimbangkan penggunaan fitur tambahan seperti warna RGB atau teknik ekstraksi fitur lain, serta model klasifikasi yang lebih kompleks seperti SVM atau CNN untuk hasil yang lebih optimal.

## REFERENSI

- [1] F. N. Cahya, R. Pebrianto, and T. A. M., "Klasifikasi Buah Segar dan Busuk Menggunakan Ekstraksi Fitur Hu-Moment, Haralick dan Histogram," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 6, no. 1, 2021, doi: 10.31294/ijcit.v6i1.10052.
- [2] M. Muhathir, M. H. Santoso, and R. Muliono, "Analysis Naïve Bayes In Classifying Fruit by Utilizing Hog Feature Extraction," *J. INFORMATICS Telecommun. Eng.*, vol. 4, no. 1, 2020, doi: 10.31289/jite.v4i1.3860.
- [3] A. Taner *et al.*, "Multiclass Apple Varieties Classification Using Machine Learning with Histogram of Oriented Gradient and Color Moments," *Appl. Sci.*, vol. 13, no. 13, 2023, doi: 10.3390/app13137682.
- [4] A. Oikonomidis, C. Catal, and A. Kassahun, "Hybrid Deep Learning-based Models for Crop Yield Prediction," *Appl. Artif. Intell.*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2022.2031823.
- [5] M. Haikal, M. Martanto, and U. Hayati, "ANALISIS SENTIMEN TERHADAP PENGGUNAAN APLIKASI GAME ONLINE PUBG MOBILE MENGGUNAKAN ALGORITMA NAIVE BAYES," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, 2024, doi: 10.36040/jati.v7i6.8174.
- [6] N. Husin, "Komparasi Algoritma Random Forest, Naïve Bayes, dan Bert Untuk Multi-Class Classification Pada Artikel Cable News Network (CNN)," *J. Esensi Infokom J. Esensi Sist. Inf. dan Sist. Komput.*, vol. 7, no. 1, 2023, doi: 10.55886/infokom.v7i1.608.
- [7] D. Saini, T. Chand, D. K. Chouhan, and M. Prakash, "A comparative analysis of automatic classification and grading methods for knee osteoarthritis focussing on X-ray images," 2021. doi: 10.1016/j.jbbe.2021.03.002.
- [8] I. Ifmalinda, A. Andasuryani, and L. Shaufana, "IDENTIFIKASI BENTUK BUAH ALPUKAT (Persea americana Mill.) DENGAN ANALISIS CITRA DIGITAL," *J. Teknol. Pertan.*, vol. 23, no. 3, 2022, doi: 10.21776/ub.jtp.2022.023.03.5.
- [9] F. A.-I. A. Putra, A. G. Sulaksono, L. T. Utomo, and A. R. Khamdani, "Klasifikasi Buah dan Sayur Menggunakan Fitur Ekstraksi HOG dan Metode KNN," *J. Inform. Polinema*, vol. 10, no. 1, 2023, doi: 10.33795/jip.v10i1.1433.
- [10] F. T. Anggraeny, B. Rahmat, and S. P. Pratama, "Deteksi Ikan Dengan Menggunakan Algoritma Histogram of Oriented Gradients," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 15, no. 2, 2020, doi: 10.30872/jim.v15i2.4648.
- [11] H. M. Ahmed, "Texture Analysis and Classification using Local Binary Patterns and Statistical Features," pp. 79–88, 2024.
- [12] M. Mukhopadhyay, A. Dey, A. Ghosh, and R. N. Shaw, "Facial Emotion Recognition Based on Textural Pattern and Histogram of Oriented Gradient," in *Communications in Computer and Information Science*, 2023. doi: 10.1007/978-3-031-25088-0\_9.
- [13] M. Afriansyah, J. Saputra, Y. Sa'adati, and Valian Yoga Pudya Ardhana, "Optimasi Algoritma Naïve Bayes Untuk Klasifikasi Buah Apel Berdasarkan Fitur Warna RGB," *Bull. Comput. Sci. Res.*, vol. 3, no. 3, pp. 242–249, 2023, doi: 10.47065/bulletincsr.v3i3.251.
- [14] R. I. Borman and M. Wati, "Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes," *J. Ilm. Fak. Ilmu Komput.*, vol. 09, no. 01, pp. 25–34, 2020.
- [15] S. Bakheet and A. Al-Hamadi, "A framework for instantaneous driver drowsiness detection based on improved HOG features and naïve bayesian classification," *Brain Sci.*, vol. 11, no. 2, pp. 1–15, 2021, doi: 10.3390/brainsci11020240.
- [16] T. Goksu, Z. Kaya, and S. Sahmoud, "Classification of Fruit Images as Fresh and Rotten Using Convolutional Neural Networks," in *2023 3rd International Conference on Computing and Information Technology, ICCIT 2023*, 2023. doi: 10.1109/ICCIT58132.2023.10273897.