JTIS : Journal of Technopreneurship and Information Systems
Vol. xx, No. x, Bulan Tahun, hal. 115~122
E-ISSN: 2614-3054; P-ISSN: 2614-3062, accredited by Kemenristekdikti, Sinta 5
DOI: 10.36085

Klasifikasi Biji Kopi Arabika dan Robusta Menggunakan Gaussian Naive Bayes Berbasis Fitur HOG

[1]Yulia Darnita, [2] Muhammad Hafiz Pebriansyah, [3]Ardi Wijaya, [4] Yuza Reswan
[1,2]Universitas Muhammadiyah Bengkulu, Indonesia
[1]yuliadarnita@umb.ac.id, [2]*bengkuluhafiz10@gmail.com*, [3]ardiwijaya@umb.ac.id, [4]yuzareswan@umb.ac.id

Article Info

Article history:
Received, 2026-06-05
Revised, 2026-07-15
Accepted, 2026-07-20

Kata Kunci:
klasifikasi_biji_kopi
Gaussian_Naive_Bayes
HOG
Arabika Robusta
Pengolahan citra

Keywords:
coffee_bean_classification
Gaussian_Naive_Bayes
HOG
Arabica_Robusta
image_processing

ABSTRAK

Klasifikasi biji kopi Arabika dan Robusta diperlukan untuk mendukung identifikasi jenis kopi yang lebih konsisten berbasis citra digital. Penelitian ini menerapkan Gaussian Naive Bayes menggunakan fitur Histogram of Oriented Gradients (HOG). Dataset terdiri atas 240 citra latih dan 80 citra uji yang terbagi seimbang antara Arabika dan Robusta. Citra diubah menjadi 64x64 piksel, dikonversi ke grayscale, kemudian diekstraksi menggunakan HOG dengan CellSize 8x8 sehingga menghasilkan 1.764 fitur. Fitur distandarkan sebelum parameter prior, mean, dan varians setiap kelas dihitung. Pada 80 data uji, model menghasilkan 63 prediksi benar dengan akurasi 78,75%. Confusion matrix menunjukkan 35 dari 40 Arabika dan 28 dari 40 Robusta terklasifikasi benar. Precision, recall, dan F1-score masing-masing adalah 0,745; 0,875; 0,805 untuk Arabika serta 0,848; 0,700; 0,767 untuk Robusta. Hasil menunjukkan bahwa Gaussian Naive Bayes dapat digunakan untuk klasifikasi dua jenis biji kopi, meskipun kesalahan masih dominan pada Robusta yang diprediksi sebagai Arabika.

ABSTRACT

Classification of Arabica and Robusta coffee beans is needed to support more consistent type identification from digital images. This study applied Gaussian Naive Bayes using Histogram of Oriented Gradients (HOG) features. The dataset contained 240 training images and 80 testing images equally distributed between Arabica and Robusta. Images were resized to 64x64 pixels, converted to grayscale, and processed using HOG with an 8x8 CellSize, producing 1,764 features. The features were standardized before class priors, means, and variances were calculated. On 80 testing images, the model produced 63 correct predictions with an accuracy of 78.75%. The confusion matrix showed that 35 of 40 Arabica images and 28 of 40 Robusta images were correctly classified. Precision, recall, and F1-score were 0.745, 0.875, and 0.805 for Arabica and 0.848, 0.700, and 0.767 for Robusta. The results show that Gaussian Naive Bayes can classify the two coffee bean types, although errors were dominated by Robusta images predicted as Arabica.

This is an open access article under the CC BY-SA license.

Penulis Korespondensi:

Muhammad Hafiz Pebriansyah,
Program Studi Teknik Informatika,
Universitas Muhammadiyah Bengkulu, Indonesia,
Email: bengkuluhafiz10@gmail.com

1. PENDAHULUAN

Kopi merupakan komoditas pertanian bernilai ekonomi tinggi dan menjadi salah satu produk penting dalam perdagangan. Indonesia dikenal sebagai salah satu negara produsen kopi utama, dengan Arabika dan Robusta sebagai dua jenis yang paling banyak diperdagangkan [1]. Perbedaan jenis berpengaruh terhadap karakteristik fisik, mutu, cita rasa, dan nilai jual, sehingga proses identifikasi biji kopi perlu dilakukan secara konsisten agar informasi jenis yang dihasilkan dapat dipertanggungjawabkan.

Identifikasi biji kopi secara manual umumnya mengandalkan pengamatan terhadap bentuk, ukuran, warna, tekstur permukaan, dan ciri visual lain. Ketika jumlah sampel meningkat, penilaian visual dapat dipengaruhi pengalaman pengamat, pencahayaan, posisi objek, dan variasi kondisi citra. Pengolahan citra digital menawarkan pendekatan yang lebih terukur karena citra dapat diubah menjadi representasi numerik, dipra-pemroses, kemudian dianalisis menggunakan metode pembelajaran mesin [13], [14].

Pada penelitian ini, karakteristik visual direpresentasikan menggunakan Histogram of Oriented Gradients (HOG). HOG membentuk deskriptor berdasarkan distribusi orientasi gradien lokal sehingga informasi bentuk dan tekstur dapat diringkas menjadi vektor fitur. Representasi berbasis fitur penting karena algoritma klasifikasi tidak bekerja langsung terhadap objek fisik, melainkan terhadap nilai numerik yang menggambarkan pola visual objek. Penelitian klasifikasi citra menunjukkan bahwa pemilihan representasi fitur sangat memengaruhi performa model [3], [16], [18].

Metode klasifikasi yang digunakan adalah Gaussian Naive Bayes. Naive Bayes merupakan pendekatan probabilistik berdasarkan Teorema Bayes dengan asumsi bahwa fitur bersifat independen bersyarat terhadap kelas. Untuk data kontinu, Gaussian Naive Bayes memodelkan distribusi setiap fitur pada masing-masing kelas menggunakan parameter mean dan varians. Keunggulan metode ini adalah proses pelatihan yang sederhana, kebutuhan komputasi relatif ringan, dan kemampuan menghasilkan probabilitas kelas secara langsung. Pada klasifikasi mutu biji kopi Robusta, pendekatan Naive Bayes telah digunakan dan menunjukkan bahwa metode probabilistik relevan untuk pengenalan karakteristik kopi [2].

Berbagai studi memperlihatkan bahwa kinerja Naive Bayes bergantung pada karakteristik data dan representasi fitur. Perbandingan dengan K-Nearest Neighbor telah dilakukan pada klasifikasi status gizi, penyakit jantung, data sosial ekonomi, stroke, daun herbal, loyalitas pelanggan, kelulusan mahasiswa, tingkat stres, dan kecelakaan lalu lintas [4]-[12]. Hasil yang berbeda antarstudi menegaskan bahwa performa algoritma tidak dapat dipisahkan dari distribusi data, hubungan antarfitur, ukuran dataset, dan strategi pra-pemrosesan.

Dalam konteks citra biji kopi, penelitian terdahulu juga menggunakan K-NN maupun CNN [3], [18], [19]. Model berbasis deep learning mampu mempelajari fitur secara otomatis, tetapi umumnya memiliki kompleksitas komputasi lebih tinggi [17]. Penelitian ini mengambil pendekatan yang lebih sederhana dengan menggabungkan HOG dan Gaussian Naive Bayes. Nilai penelitian terletak pada penerapan model probabilistik pada fitur HOG 1.764 dimensi untuk membedakan Arabika dan Robusta, disertai evaluasi confusion matrix serta metrik per kelas agar pola kesalahan dapat dianalisis secara rinci.

Tujuan penelitian adalah membangun proses klasifikasi citra biji kopi Arabika dan Robusta menggunakan Gaussian Naive Bayes berbasis fitur HOG, mengukur kinerja model pada data uji yang terpisah, dan menganalisis kesalahan klasifikasi tiap kelas. Hasil penelitian diharapkan memberikan gambaran empiris mengenai kemampuan serta keterbatasan Gaussian Naive Bayes ketika diterapkan pada fitur citra berukuran tinggi dan dapat menjadi dasar pengembangan sistem klasifikasi kopi yang lebih robust.

2. METODE PENELITIAN

Penelitian menggunakan desain eksperimen kuantitatif. Proses dimulai dari penyiapan dataset citra biji kopi, pra-pemrosesan, ekstraksi fitur HOG, standardisasi fitur, pembentukan model Gaussian Naive Bayes, prediksi data uji, dan evaluasi hasil. Seluruh data uji dipisahkan dari data latih sehingga metrik yang diperoleh menggambarkan kemampuan model pada citra yang tidak digunakan ketika parameter kelas dihitung.

Dataset terdiri atas 240 citra latih dan 80 citra uji. Data latih terbagi menjadi 120 citra Arabika dan 120 citra Robusta, sedangkan data uji terdiri atas 40 Arabika dan 40 Robusta. Komposisi yang seimbang memudahkan interpretasi metrik karena setiap kelas memiliki jumlah sampel yang sama. Label kelas diperoleh berdasarkan struktur folder dataset yang memisahkan citra Arabika dan Robusta.

Pra-pemrosesan dilakukan dengan mengubah ukuran setiap citra menjadi 64x64 piksel dan mengonversi citra RGB menjadi grayscale. Penyeragaman ukuran memastikan seluruh sampel memiliki dimensi yang identik sebelum ekstraksi fitur, sedangkan grayscale mengurangi kompleksitas kanal warna dan memusatkan pemrosesan pada distribusi intensitas. Informasi mean grayscale, ukuran citra, jumlah kanal, serta panjang fitur juga dicatat sebagai bagian dari validasi proses.

Ekstraksi fitur dilakukan menggunakan Histogram of Oriented Gradients dengan CellSize 8x8. Pada citra 64x64 piksel, konfigurasi tersebut menghasilkan vektor HOG sepanjang 1.764 elemen. HOG digunakan untuk menonjolkan perubahan gradien lokal yang berkaitan dengan bentuk, tepi, dan tekstur. Dengan demikian, setiap citra diubah dari matriks intensitas menjadi satu vektor numerik berdimensi tetap yang dapat digunakan oleh model klasifikasi.

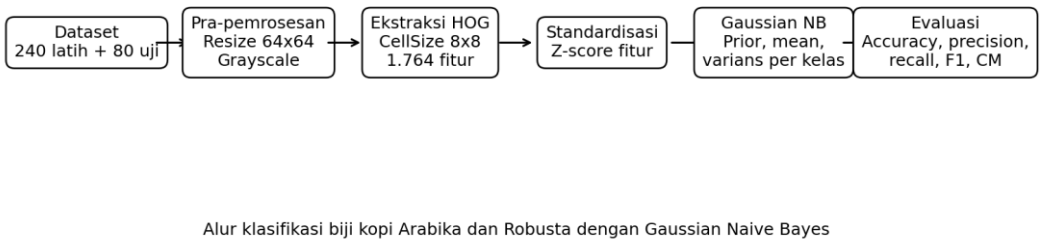
Sebelum pembentukan model, vektor fitur distandarkan menggunakan z-score. Standardisasi menggeser setiap fitur terhadap rata-rata data latih dan membaginya dengan simpangan baku sehingga skala antarf fitur menjadi lebih seragam. Parameter standardisasi dari data latih kemudian digunakan secara konsisten ketika memproses data uji agar tidak terjadi kebocoran informasi dari data pengujian.

Gaussian Naive Bayes dibentuk dengan menghitung prior kelas, mean, dan varians setiap fitur pada kelas Arabika dan Robusta. Karena jumlah data latih kedua kelas seimbang, prior empiris masing-masing kelas adalah setara. Pada tahap prediksi, log-posterior dihitung untuk setiap kelas dari log prior dan akumulasi log-likelihood Gaussian seluruh fitur. Label dengan nilai log-posterior terbesar dipilih sebagai hasil prediksi. Penggunaan log-posterior menghindari masalah numerik akibat perkalian probabilitas yang sangat kecil pada fitur berdimensi tinggi.

Evaluasi dilakukan pada 80 data uji menggunakan accuracy, confusion matrix, precision, recall, dan F1-score. Accuracy mengukur proporsi seluruh prediksi yang benar. Precision menunjukkan ketepatan prediksi suatu kelas, sedangkan recall mengukur kemampuan model menemukan seluruh sampel yang benar-benar termasuk kelas tersebut. F1-score digunakan untuk merangkum keseimbangan precision dan recall. Confusion matrix digunakan untuk melihat arah kesalahan Arabika menjadi Robusta maupun Robusta menjadi Arabika.

Sistem diimplementasikan pada MATLAB R2022a. Proses pemuatan citra, pra-pemrosesan, ekstraksi HOG, standardisasi, perhitungan parameter Gaussian Naive Bayes, prediksi, evaluasi, dan ekspor keluaran diintegrasikan dalam antarmuka eksperimen. Penyimpanan keluaran per citra memungkinkan hasil prediksi ditelusuri kembali sehingga perhitungan manual terhadap sampel tertentu dapat diverifikasi.

Alur penelitian ditunjukkan pada Gambar 1. Setiap tahap menggunakan hasil dari tahap sebelumnya secara berurutan: dataset menjadi citra terstandar, citra dikonversi menjadi fitur HOG, fitur distandarkan, parameter probabilistik dihitung dari data latih, kemudian data uji diklasifikasikan dan dievaluasi. Struktur ini memastikan bahwa pembentukan model Gaussian Naive Bayes dan pengukuran performa dapat direplikasi dengan parameter yang sama.



Gambar 1. Alur penelitian klasifikasi biji kopi menggunakan Gaussian Naive Bayes

Tabel 1. Konfigurasi dataset dan parameter Gaussian Naive Bayes

Komponen	Konfigurasi
Data latih	240 citra: 120 Arabika dan 120 Robusta
Data uji	80 citra: 40 Arabika dan 40 Robusta
Ukuran citra	64x64 piksel
Pra-pemrosesan	Resize dan grayscale
Ekstraksi fitur	HOG, CellSize 8x8
Panjang fitur	1.764 elemen
Standardisasi fitur	Z-score menggunakan parameter data latih
Klasifikasi	Gaussian Naive Bayes: prior, mean, varians
Platform	MATLAB R2022a

Contoh citra Arabika dan Robusta ditampilkan pada Gambar 2. Variasi orientasi, ukuran objek, pencahayaan, dan tekstur membuat tahap standarisasi serta ekstraksi HOG diperlukan agar setiap citra dapat direpresentasikan dalam bentuk numerik yang konsisten. Model kemudian mempelajari distribusi fitur pada masing-masing kelas, bukan menggunakan ciri visual yang dinilai secara subjektif oleh pengamat.



Gambar 2. Contoh citra Arabika dan Robusta pada dataset penelitian

Tabel 2. Metrik evaluasi Gaussian Naive Bayes

Metrik	Tujuan	Interpretasi
Accuracy	Proporsi prediksi benar	Kinerja keseluruhan
Precision	Ketepatan prediksi positif kelas	Rendah jika banyak false positive
Recall	Kemampuan menemukan sampel kelas	Rendah jika banyak false negative
F1-score	Rata-rata harmonik precision-recall	Keseimbangan performa kelas
Confusion matrix	Distribusi prediksi terhadap label asli	Arah dan jumlah kesalahan

3. HASIL DAN ANALISIS

Pengujian dilakukan terhadap 80 citra yang terdiri atas 40 Arabika dan 40 Robusta. Setiap citra melalui resize 64x64, konversi grayscale, ekstraksi HOG 1.764 fitur, standardisasi menggunakan parameter data latih, dan perhitungan log-posterior Gaussian Naive Bayes. Dari seluruh data uji, 63 citra diprediksi benar dan 17 citra salah sehingga akurasi keseluruhan mencapai 78,75%.

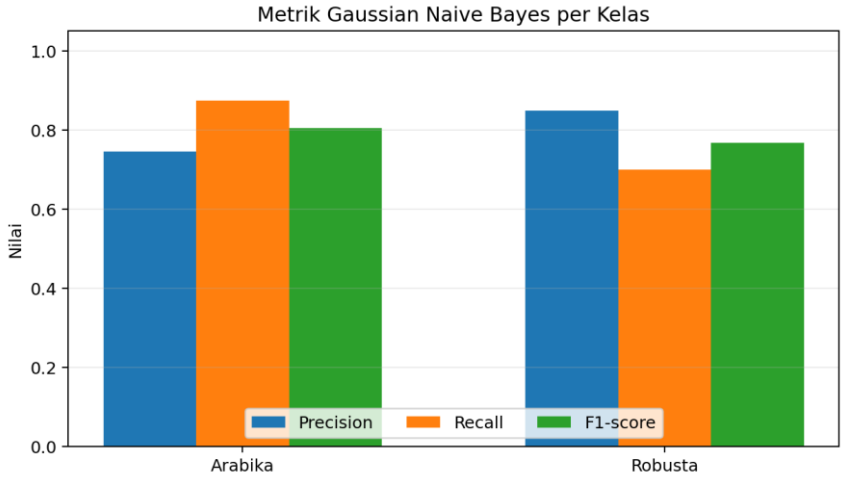
Nilai akurasi tersebut menunjukkan bahwa model telah mampu membedakan dua jenis biji kopi pada sebagian besar sampel, tetapi masih terdapat ruang perbaikan. Dengan distribusi kelas yang seimbang, akurasi 78,75% dapat langsung dibaca sebagai 63 prediksi benar dari 80 citra. Kesalahan sebesar 21,25% tidak tersebar merata antar kelas, sehingga analisis confusion matrix diperlukan untuk mengetahui kelas mana yang lebih sulit dikenali.

Kinerja Gaussian Naive Bayes perlu dipahami sesuai karakteristik metodenya. Model mengestimasi distribusi tiap fitur secara terpisah pada setiap kelas dan menggabungkannya berdasarkan asumsi independensi bersyarat. Fitur HOG dihasilkan dari gradien lokal yang secara praktis dapat memiliki korelasi antar elemen. Ketika asumsi independensi tidak sepenuhnya sesuai dengan struktur fitur, probabilitas gabungan dapat menjadi kurang optimal meskipun fitur individual tetap mengandung informasi diskriminatif. Hal ini menjadi salah satu penjelasan konseptual terhadap kesalahan yang masih muncul.

Tabel 3 merangkum performa keseluruhan model. Selain akurasi, jumlah prediksi salah digunakan untuk menegaskan besar kesalahan yang masih perlu dianalisis. Model menghasilkan 63 prediksi benar dan 17 prediksi salah. Pada penelitian ini angka tersebut dijadikan hasil utama karena diperoleh dari konfigurasi final dataset, HOG, standardisasi, dan Gaussian Naive Bayes yang digunakan pada proses pengujian.

Tabel 3. Ringkasan hasil pengujian Gaussian Naive Bayes

Indikator	Nilai	Jumlah	Keterangan
Prediksi benar	78,75%	63 dari 80	Hasil klasifikasi sesuai label asli
Prediksi salah	21,25%	17 dari 80	Hasil klasifikasi tidak sesuai label asli



Gambar 3. Precision, recall, dan F1-score Gaussian Naive Bayes per kelas

Distribusi hasil prediksi menunjukkan perbedaan performa antar kelas. Dari 40 citra Arabika, sebanyak 35 citra berhasil dikenali sebagai Arabika dan lima citra diprediksi sebagai Robusta. Dengan demikian, recall Arabika adalah 0,875. Nilai ini menunjukkan bahwa model relatif baik dalam menangkap sebagian besar sampel Arabika pada data uji.

Pada kelas Robusta, 28 dari 40 citra diprediksi benar sebagai Robusta, sedangkan 12 citra diprediksi sebagai Arabika. Recall Robusta sebesar 0,700 lebih rendah daripada recall Arabika. Selisih tersebut menunjukkan bahwa kesalahan model lebih banyak berasal dari Robusta yang memiliki pola HOG lebih dekat dengan distribusi fitur Arabika menurut estimasi Gaussian Naive Bayes.

Confusion matrix pada Tabel 5 dan Gambar 4 memperlihatkan pola kesalahan secara langsung. Kesalahan Arabika menjadi Robusta berjumlah lima, sedangkan kesalahan Robusta menjadi Arabika berjumlah dua belas. Pola yang tidak simetris ini penting karena akurasi keseluruhan saja tidak menjelaskan bahwa kemampuan model mengenali Robusta lebih rendah. Oleh sebab itu, metrik per kelas digunakan untuk melengkapi evaluasi.

Tabel 4. Distribusi hasil prediksi per kelas

Kelas	Data uji	Prediksi benar
Arabika	40	35
Robusta	40	28

Tabel 5. Confusion matrix Gaussian Naive Bayes

Asli / Prediksi	Arabika	Robusta
Arabika	35	5
Robusta	12	28

Metrik per kelas pada Tabel 6 menunjukkan precision Arabika sebesar 0,744681, recall 0,875, dan F1-score 0,804598. Precision Arabika lebih rendah daripada recall karena terdapat 12 citra Robusta yang ikut diprediksi sebagai Arabika. Dengan kata lain, model cukup sensitif menemukan Arabika, tetapi prediksi Arabika masih mengandung false positive dari kelas Robusta.

Tabel 6. Precision, recall, dan F1-score Gaussian Naive Bayes

Kelas	Precision	Recall	F1-score	Support
Arabika	0,744681	0,875	0,804598	40
Robusta	0,848485	0,700	0,767123	40
Macro avg	0,796583	0,787500	0,785861	80
Weighted avg	0,796583	0,787500	0,785861	80

Untuk Robusta, precision mencapai 0,848485, recall 0,700, dan F1-score 0,767123. Precision yang lebih tinggi menunjukkan bahwa ketika model memutuskan suatu citra sebagai Robusta, keputusan tersebut relatif tepat. Namun, recall yang lebih rendah menunjukkan banyak citra Robusta tidak berhasil ditemukan karena dialihkan ke Arabika. Perbedaan precision dan recall ini menegaskan bahwa interpretasi model sebaiknya tidak hanya berdasarkan akurasi.

Validasi manual dilakukan terhadap salah satu sampel, yaitu Arabika_Test_001.JPG. Setelah resize 64x64 dan grayscale, citra memiliki 4.096 piksel dengan mean grayscale 177,446533 dan panjang fitur HOG 1.764. Informasi tersebut digunakan untuk memastikan bahwa sampel uji melalui tahapan pra-pemrosesan yang sama dengan data lain sebelum nilai log-posterior dihitung.

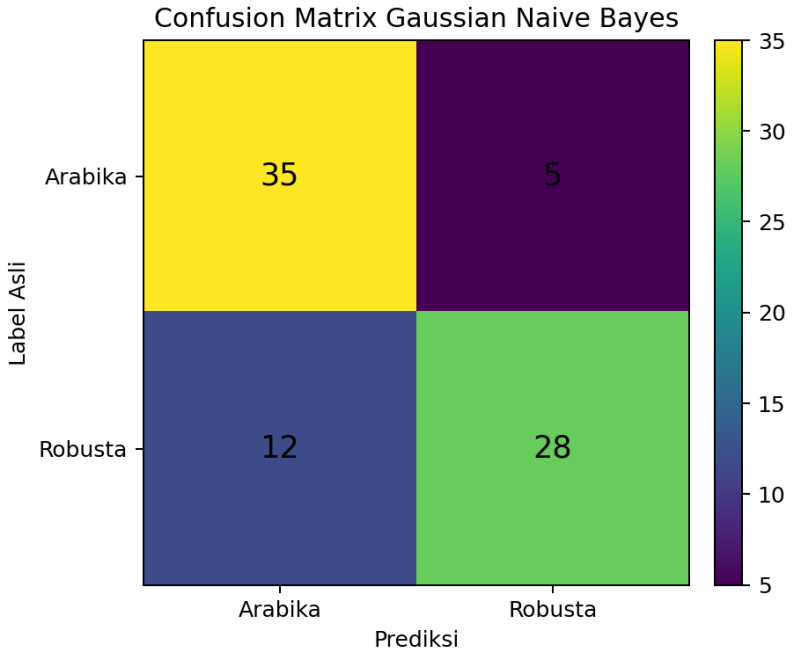
Pada sampel tersebut, log-posterior kelas Arabika diperoleh sekitar -2045,29, sedangkan log-posterior kelas Robusta sekitar -2411,93. Karena -2045,29 lebih besar daripada -2411,93, Gaussian Naive Bayes menetapkan prediksi Arabika. Prediksi tersebut sesuai dengan label asli. Contoh ini memperlihatkan mekanisme keputusan model: kelas dipilih bukan dari satu fitur tunggal, tetapi dari hasil agregasi likelihood seluruh fitur bersama prior kelas.

Komponen log-posterior juga menunjukkan besarnya kontribusi kumulatif fitur berdimensi tinggi. Nilai logPrior kedua kelas sama, yaitu sekitar -0,693147, sesuai distribusi data latih yang seimbang. Perbedaan keputusan terutama berasal dari constTerm dan quadTerm yang dihitung dari mean, varians, dan jarak fitur sampel terhadap distribusi Gaussian masing-masing kelas. Penggunaan log membantu menjaga stabilitas perhitungan ketika banyak likelihood kecil harus digabungkan.

Hasil penelitian konsisten dengan prinsip dasar bahwa Naive Bayes bekerja efektif ketika distribusi fitur per kelas dapat dimodelkan dengan baik dan ketergantungan antarf fitur tidak terlalu dominan [4]-[12]. Pada fitur HOG, setiap elemen merepresentasikan histogram gradien pada area lokal sehingga hubungan spasial antarf fitur tetap mungkin terjadi. Oleh karena itu, hasil 78,75% dapat dipahami sebagai kombinasi antara informasi diskriminatif HOG dan keterbatasan asumsi independensi Gaussian Naive Bayes.

Dibandingkan metode dengan kompleksitas lebih tinggi seperti convolutional neural network, Gaussian Naive Bayes memiliki proses pelatihan yang lebih sederhana dan kebutuhan komputasi lebih ringan [17], [19]. Keuntungan ini relevan untuk eksperimen dengan dataset menengah atau ketika interpretasi probabilistik dan efisiensi implementasi lebih diprioritaskan. Di sisi lain, model tidak mempelajari hubungan spasial citra secara langsung, sehingga kualitas representasi fitur HOG menjadi sangat menentukan.

Analisis kelas memperlihatkan bahwa peningkatan performa sebaiknya difokuskan pada Robusta. Sebanyak 30% citra Robusta tidak dikenali dengan benar, sedangkan kesalahan pada Arabika sebesar 12,5%. Beberapa kemungkinan penyebab adalah kemiripan bentuk dan tekstur antarjenis, variasi pencahayaan, sudut pengambilan, serta korelasi fitur HOG yang tidak dimodelkan secara eksplisit oleh Naive Bayes. Penambahan variasi data dan optimasi parameter HOG dapat membantu menguji hipotesis tersebut.



Gambar 4. Confusion matrix Gaussian Naive Bayes

Tabel 7. Ringkasan pra-pemrosesan sampel Arabika_Test_001.JPG

Komponen	Nilai	Keterangan	Output
File	Arabika_Test_001.JPG	Data uji	Arabika
Ukuran	64x64	Setelah resize	4.096 piksel
Mean grayscale	177,446533	Rata-rata intensitas	-
Panjang HOG	1.764	CellSize 8x8	Vektor fitur
Jumlah channel	3 -> 1	RGB ke grayscale	Siap klasifikasi

Tabel 7 merangkum pra-pemrosesan sampel Arabika_Test_001.JPG. Ukuran 64x64, jumlah 4.096 piksel, mean grayscale 177,446533, panjang HOG 1.764, dan informasi kanal menunjukkan bahwa data uji telah ditransformasikan ke format yang sesuai dengan pipeline klasifikasi sebelum masuk ke perhitungan probabilitas.

Tabel 8. Komponen log-posterior Gaussian Naive Bayes

Kelas	logPrior	constTerm	quadTerm	logPosterior
Arabika	-0,693147	-1450,73	-593,858	-2045,29
Robusta	-0,693147	-1589,29	-821,943	-2411,93

Pola kesalahan pada Tabel 9 memperlihatkan bahwa kelas Arabika menghasilkan 35 prediksi benar dan lima kesalahan, sedangkan Robusta menghasilkan 28 prediksi benar dan dua belas kesalahan. Kesalahan Robusta menjadi Arabika merupakan sumber kesalahan terbesar. Secara praktis, model cenderung lebih sering memasukkan citra yang meragukan ke kelas Arabika daripada ke Robusta.

Dari sisi precision, Robusta memperoleh nilai lebih tinggi daripada Arabika karena jumlah false positive Robusta hanya lima. Sebaliknya, precision Arabika turun akibat 12 Robusta diprediksi sebagai Arabika. Dari sisi recall, Arabika lebih unggul karena hanya lima sampel Arabika terlewat. Perbedaan ini memperlihatkan trade-off antarmetrik yang perlu dipertimbangkan apabila sistem nantinya digunakan pada proses identifikasi yang memiliki konsekuensi berbeda untuk masing-masing jenis kopi.

Tabel 9. Pola kesalahan klasifikasi Gaussian Naive Bayes

Kelas / Indikator	Benar	Salah	Recall / Nilai	Pola / Keterangan
Arabika	35	5	0,875	Arabika -> Robusta
Robusta	28	12	0,700	Robusta -> Arabika
Total	63	17	Akurasi 0,7875	80 data uji
Kesalahan dominan	-	12	30% Robusta	Robusta diprediksi Arabika

Validasi manual pada log-posterior mendukung hasil implementasi. Model menghitung prior yang sama untuk kedua kelas karena data latih seimbang, lalu membandingkan kecocokan fitur sampel terhadap distribusi Gaussian masing-masing kelas. Mekanisme ini sederhana dan transparan sehingga keputusan sebuah sampel dapat ditelusuri kembali melalui komponen perhitungan, salah satu kelebihan penting dibandingkan model yang lebih sulit diinterpretasikan.

Posisi penelitian terhadap studi terdahulu dirangkum pada Tabel 10. Studi Fata dan Avianto [2] menunjukkan penggunaan Naive Bayes pada biji kopi Robusta, sedangkan penelitian lain menggunakan K-NN atau CNN pada pengenalan biji kopi [3], [18], [19]. Penelitian ini memperluas konteks penggunaan Naive Bayes dengan mengklasifikasikan Arabika dan Robusta menggunakan fitur HOG 1.764 dimensi serta menyediakan evaluasi per kelas.

Tabel 10. Posisi penelitian terhadap beberapa studi terdahulu

Studi	Objek / Fitur	Metode	Temuan yang relevan
Fata dan Avianto [2]	Mutu biji kopi Robusta	Naive Bayes	Metode probabilistik dapat diterapkan pada klasifikasi kopi
Loka dan Marsal [4]	Status gizi	K-NN vs Naive Bayes	Performa bergantung pada karakteristik dan distribusi data
Iswara et al. [7]	Klasifikasi stroke	K-NN vs Naive Bayes	Perbandingan algoritma dipengaruhi karakteristik fitur
Sari dan Nugroho [19]	Jenis biji kopi	CNN	Model kompleks dapat mempelajari fitur citra secara otomatis
Penelitian ini	HOG 1.764 fitur	Gaussian Naive Bayes	Akurasi 78,75%; kesalahan dominan Robusta -> Arabika

Keterbatasan penelitian berada pada penggunaan satu konfigurasi utama HOG dan satu pembagian data latih-uji. Gaussian Naive Bayes juga menggunakan asumsi distribusi Gaussian per fitur dan independensi bersyarat yang belum tentu sepenuhnya sesuai dengan struktur HOG. Selain itu, variasi pencahayaan, orientasi, kamera, dan kondisi fisik biji dapat memengaruhi distribusi fitur. Karena itu, akurasi 78,75% tidak boleh dipandang sebagai nilai tetap untuk seluruh kondisi klasifikasi biji kopi.

Penelitian selanjutnya dapat menguji variasi CellSize dan konfigurasi blok HOG, menerapkan validasi silang, melakukan seleksi fitur untuk mengurangi korelasi atau dimensi yang kurang informatif, serta memperbanyak dataset dengan variasi kondisi pengambilan citra. Perbandingan lanjutan dengan model probabilistik lain, SVM, Random Forest, maupun metode deep learning dapat dilakukan untuk mengetahui trade-off antara akurasi, kompleksitas, kebutuhan data, dan interpretabilitas.

Secara keseluruhan, kombinasi HOG dan Gaussian Naive Bayes mampu melakukan klasifikasi Arabika dan Robusta dengan akurasi 78,75%. Model lebih baik dalam menemukan Arabika daripada Robusta berdasarkan recall, sementara precision Robusta lebih tinggi. Temuan ini menunjukkan bahwa model sudah dapat digunakan sebagai baseline probabilistik berbasis fitur, tetapi peningkatan representasi dan strategi pemodelan masih diperlukan untuk mengurangi kesalahan Robusta menjadi Arabika.

4. KESIMPULAN

Penelitian berhasil menerapkan Gaussian Naive Bayes untuk klasifikasi citra biji kopi Arabika dan Robusta menggunakan pra-pemrosesan 64x64 piksel dan grayscale, ekstraksi HOG CellSize 8x8 yang menghasilkan 1.764 fitur, serta standardisasi fitur sebelum estimasi parameter probabilistik. Pada 80 data uji, model menghasilkan 63 prediksi benar dengan akurasi 78,75%. Kelas Arabika memperoleh precision 0,744681, recall 0,875, dan F1-score 0,804598, sedangkan Robusta memperoleh precision 0,848485, recall 0,700, dan F1-score 0,767123. Confusion matrix menunjukkan kesalahan terbesar berupa 12 citra Robusta yang diprediksi sebagai Arabika. Dengan demikian, Gaussian Naive Bayes dapat digunakan sebagai model klasifikasi berbasis fitur HOG yang sederhana dan interpretatif, tetapi pengembangan selanjutnya perlu diarahkan pada perbaikan representasi fitur, seleksi fitur, optimasi HOG, validasi silang, dan penambahan variasi dataset untuk meningkatkan kemampuan generalisasi.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Bengkulu, serta pihak-pihak yang telah mendukung proses penelitian dan penyusunan skripsi.

REFERENSI

[1] FTNews, “Indonesia Jadi Negara Penghasil Kopi Terbesar Ketiga di Dunia”, 2023.

[2] N. Fata and D. Avianto, “Klasifikasi Mutu Biji Kopi Robusta Menggunakan Algoritma Naive Bayes”, *Jurnal Ilmiah Teknologi Pertanian dan Komputer*, vol. 9, no. 1, pp. 33-41, 2024.

[3] A. Fadjeri, “Klasifikasi bentuk biji kopi menggunakan metode K-Nearest Neighbor (K-NN) berbasis fitur morfologis”, *Jurnal Informatika*, vol. 10, no. 2, pp. 115-123, 2023.

[4] A. Loka and E. Marsal, “Perbandingan Algoritma K-NN dan Naive Bayes dalam Klasifikasi Status Gizi Balita”, *Jurnal Informatika dan Sistem Cerdas*, vol. 7, no. 3, pp. 152-160, 2023.

[5] N. Sahar, “Analisis Kinerja Naive Bayes dan K-NN dalam Klasifikasi Penyakit Jantung”, *Jurnal Ilmu Komputer dan Aplikasi*, vol. 14, no. 1, pp. 22-30, 2020.

[6] R. Pemungkas and M. Asnawi, “Analisis Naive Bayes dan K-NN dalam Klasifikasi Keluarga Miskin Berdasarkan Data Sosial Ekonomi”, *Jurnal Data Mining Indonesia*, vol. 6, no. 2, pp. 98-106, 2024.

[7] I. B. A. I. Iswara, I. B. G. Anandita, and M. Dahul, “Comparative analysis of Naïve Bayes and K-Nearest Neighbor (KNN) algorithms in stroke classification”, *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 6, no. 3, pp. 1415-1424, 2024.

[8] B. I. Nugroho, M. W. Khusni, P. S. Ananda, and G. Gunawan, “Comparison of Naïve Bayes and KNN for herbal leaf classification”, *Jurnal Mandiri IT*, vol. 13, no. 1, pp. 18-27, 2024.

[9] A. S. Ramadhan, L. Magdalena, and M. Febima, “Perbandingan kinerja algoritma K-Nearest Neighbor dan Naive Bayes untuk klasifikasi loyalitas pelanggan”, *RIGGS: Journal of Artificial Intelligence and Digital Business*, 2025.

[10] Y. N. Rumbia et al., “Perbandingan metode KNN dan Naive Bayes untuk klasifikasi kelulusan mahasiswa”, *Jurnal PTI*, vol. 12, no. 1, pp. 7-13, 2025.

[11] Y. C. Tapidingan and D. Paseru, “Comparative analysis of classification methods of KNN and Naïve Bayes to determine stress level”, *Indonesian Journal of Information Systems*, vol. 2, no. 2, 2023.

[12] D. Wisdayani, “Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Menggunakan Algoritma K-NN dan Naive Bayes”, *Jurnal Rekayasa Sistem dan Informatika*, vol. 5, no. 4, pp. 187-196, 2019.

[13] M. Bachtiar, *Pengolahan Citra Digital: Teori dan Implementasi*. Andi Publisher

[14] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques, 3rd ed.*. Morgan Kaufmann

[15] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann

[16] E. Budianita, A. Rahman, and B. Santoso, “Analisis Performa Algoritma K-Nearest Neighbor pada Klasifikasi Data Citra”, *Jurnal Teknologi Informasi dan Sains*, vol. 8, no. 2, pp. 115-124, 2024.

[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press

[18] D. Putra and J. Santoso, “Penerapan Metode KNN untuk Klasifikasi Mutu Biji Kopi Arabika dan Robusta”, *Jurnal Teknologi Pertanian Indonesia*, vol. 12, no. 2, pp. 45-52, 2021.

[19] M. Sari and A. Nugroho, “Pengenalan Jenis Biji Kopi Menggunakan CNN”, *Jurnal Informatika dan Sains Komputer*, vol. 14, no. 3, pp. 210-218, 2022.