

ABSTRAK

PERBANDINGAN AKURASI KLASIFIKASI PADA BIJI KOPI MENGUNAKAN METODE K-NN DAN NAIVE BAYES

Nama : Muhammad Hafiz Pebriansyah
NPM : 2155201201
Pembimbing : Yulia Darnita, S.Kom, M.Kom

Klasifikasi biji kopi Arabika dan Robusta merupakan salah satu permasalahan penting dalam bidang pengolahan citra digital dan pembelajaran mesin, mengingat perbedaan jenis kopi berpengaruh terhadap kualitas dan nilai jual. Penelitian ini bertujuan untuk membandingkan tingkat akurasi metode K-Nearest Neighbor (KNN) dan Naïve Bayes dalam mengklasifikasikan biji kopi Arabika dan Robusta berdasarkan citra digital. Dataset yang digunakan terdiri dari 240 data latih dan 80 data uji, yang masing-masing terbagi secara seimbang antara kelas Arabika dan Robusta. Tahapan penelitian meliputi pra-pemrosesan citra berupa *resize*, konversi *grayscale*, dan normalisasi, dilanjutkan dengan ekstraksi fitur menggunakan metode *Histogram of Oriented Gradients* (HOG). Proses klasifikasi dilakukan menggunakan metode KNN dan Naïve Bayes, kemudian dievaluasi menggunakan confusion matrix dan perhitungan akurasi. Hasil pengujian menunjukkan bahwa metode KNN mampu mengklasifikasikan 68 dari 80 data uji dengan tingkat akurasi sebesar 85%, sedangkan metode Naïve Bayes mengklasifikasikan 63 data dengan tingkat akurasi sebesar 78,75%. Perbedaan performa ini disebabkan oleh karakteristik metode KNN yang lebih sesuai untuk fitur HOG yang memiliki keterkaitan antar fitur, dibandingkan dengan asumsi independensi fitur pada Naïve Bayes. Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode K-Nearest Neighbor memiliki performa yang lebih baik dibandingkan Naïve Bayes dalam klasifikasi biji kopi Arabika dan Robusta pada penelitian ini.

Kata Kunci: Klasifikasi Biji Kopi, Pengolahan Citra Digital, K-Nearest Neighbor, Naïve Bayes, *Histogram of Oriented Gradients*.

ABSTRACT

COMPARASION OF CLASSIFICATION ACCURACY OF COFFEE BEANS USING K-NN AND NAIVE BAYES

Name : Muhammad Hafiz Pebriansyah
Student ID : 2155201201
Supervisor : Yulia Darnita, S.Kom, M.Kom

The classification of Arabica and Robusta coffee beans is an important issue in the field of digital image processing and machine learning, given that differences in coffee types affect quality and selling value. This study aims to compare the accuracy levels of the K-Nearest Neighbor (KNN) and Naïve Bayes methods in classifying Arabica and Robusta coffee beans based on digital images. The dataset used consisted of 240 training data and 80 test data, which were divided equally between the Arabica and Robusta classes. The research stages included image pre-processing in the form of resizing, grayscale conversion, and normalization, followed by feature extraction using the Histogram of Oriented Gradients (HOG) method. The classification process was carried out using the KNN and Naïve Bayes methods, then evaluated using a confusion matrix and accuracy calculations. The test results showed that the KNN method was able to classify 68 of the 80 test data with an accuracy rate of 85%, while the Naïve Bayes method classified 63 data with an accuracy rate of 78.75%. This difference in performance is due to the characteristics of the KNN method, which is more suitable for HOG features that have inter-feature correlations, compared to the assumption of feature independence in Naïve Bayes. Based on the results of this study, it can be concluded that the K-Nearest Neighbor method performs better than Naïve Bayes in classifying Arabica and Robusta coffee beans in this study.

Keywords: *Coffee Bean Classification, Digital Image Processing, K-Nearest Neighbor, Naïve Bayes, Histogram of Oriented Gradients.*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kopi Indonesia terus mempertahankan posisinya sebagai salah satu komoditas unggulan di pasar global. Data menunjukkan bahwa Indonesia merupakan negara penghasil kopi terbesar ketiga di dunia, dengan total produksi mencapai sekitar 794,8 ribu ton pada tahun 2022 (FTNews, 2023). Dalam produksi ini, mayoritas adalah kopi Robusta, dan sisanya Arabika. Kopi khas Bengkulu, yang memiliki karakteristik cita rasa dan bentuk biji yang unik, menjadi bagian penting dari komoditas ini. Dalam konteks ilmiah, klasifikasi biji kopi, khususnya untuk membedakan jenis Robusta dan Arabika, masih menjadi tantangan yang menarik dalam bidang pengolahan citra dan pembelajaran mesin (Fata dan Avianto, 2024).

Proses klasifikasi biji kopi yang masih banyak dilakukan secara manual menimbulkan ketidakseragaman data dan kesulitan dalam standarisasi penelitian. Oleh karena itu, diperlukan eksplorasi potensi digitalisasi proses klasifikasi menggunakan pendekatan teknologi berbasis kecerdasan buatan. Permasalahan mendasar dalam studi ini adalah kebutuhan untuk membandingkan kinerja metode klasifikasi untuk mencapai hasil yang konsisten dan berbasis data kuantitatif.

Penelitian ini penting karena menawarkan kajian ilmiah mengenai komparasi metode klasifikasi citra digital. Studi literatur menunjukkan bahwa algoritma seperti K-NN dan Naive Bayes telah diterapkan secara luas dalam berbagai bidang klasifikasi, termasuk klasifikasi data sosial maupun kualitas udara, namun perbandingan performa keduanya pada citra biji kopi masih perlu dikaji secara mendalam (Pemungkas dan Asnawi, 2024; Budianita et al., 2024). Fokus utama adalah mengevaluasi kinerja dua algoritma ini dalam mengidentifikasi jenis biji kopi Arabika dan Robusta.

Perbandingan ini akan menunjukkan metode mana yang memiliki performa akurasi tertinggi secara akademis untuk klasifikasi biji kopi berdasarkan ciri visualnya.

1.2 Pertanyaan Penelitian

Permasalahan utama yang diangkat dalam penelitian ini berfokus pada perbandingan akurasi dua metode klasifikasi populer dalam pembelajaran mesin, yaitu K-Nearest Neighbor (K-NN) dan Naive Bayes, untuk mengidentifikasi jenis biji kopi berdasarkan citra digital. Penelitian ini juga mengevaluasi fitur citra mana yang paling berpengaruh terhadap tingkat akurasi klasifikasi.

Dengan demikian, pertanyaan penelitiannya adalah:

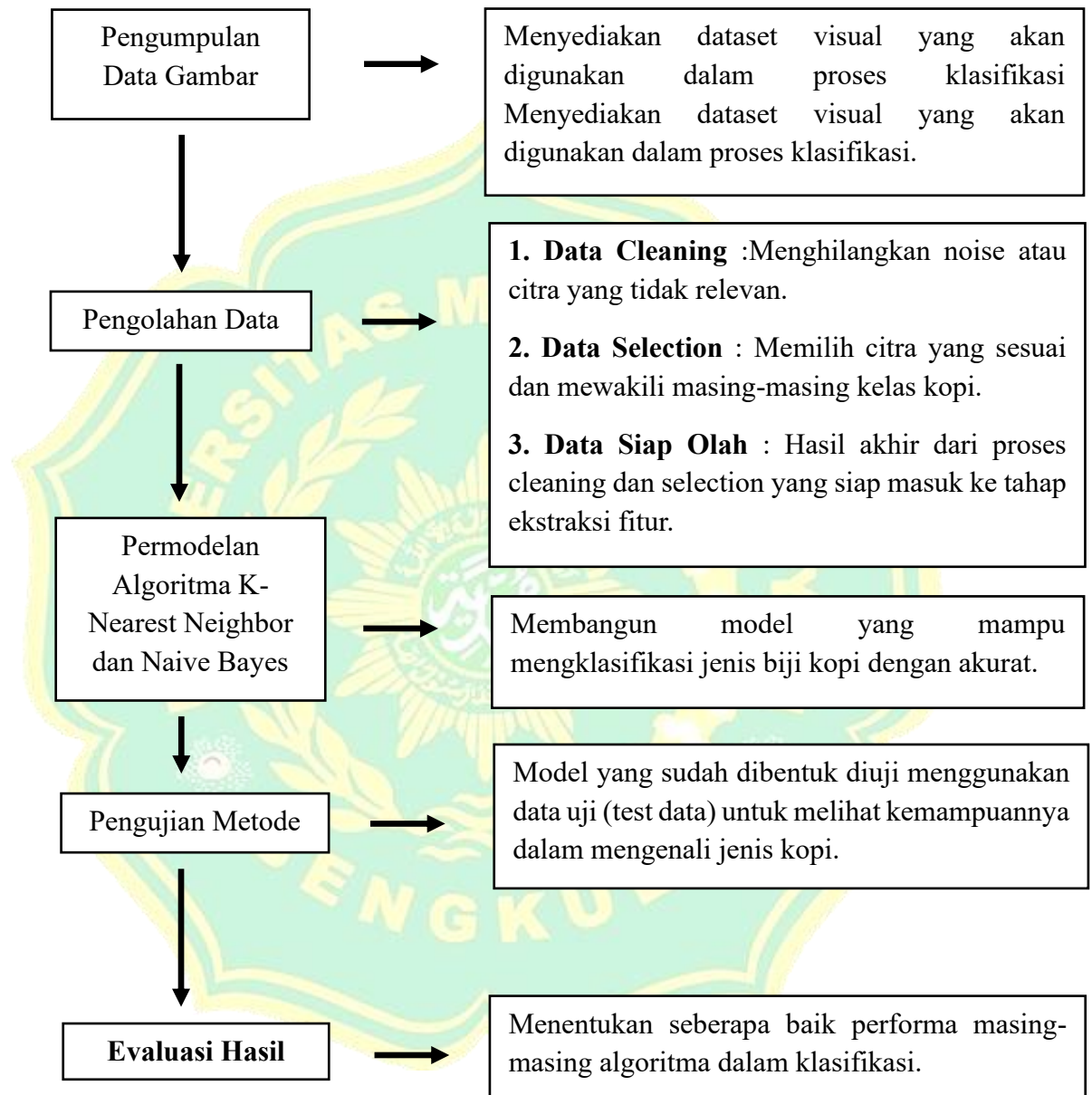
1. Berapa perbandingan tingkat akurasi antara metode K-Nearest Neighbor (K-NN) dan Naive Bayes dalam mengklasifikasikan jenis biji kopi (Arabika dan Robusta) berdasarkan citra digital?
2. Fitur citra (warna, tekstur, dan bentuk) manakah yang paling berpengaruh terhadap tingkat akurasi klasifikasi biji kopi menggunakan kedua metode tersebut?
3. Apa saja tahapan utama dalam menggunakan K-Nearest Neighbor dan Naive Bayes, bagaimana masing-masing tahapan berkontribusi dalam klasifikasi biji kopi?

1.3 Tujuan Penelitian

1. Membandingkan secara komprehensif performa akademis dan tingkat akurasi dua algoritma pembelajaran mesin, yaitu K-Nearest Neighbor (K-NN) dan Naive Bayes, dalam mengklasifikasikan jenis biji kopi Arabika dan Robusta menggunakan data citra digital.
2. Mengevaluasi dan mengidentifikasi fitur citra yang paling optimal dan memiliki kontribusi paling signifikan, seperti warna, tekstur, atau bentuk, terhadap akurasi model klasifikasi.
3. Hasil akhir dari penelitian ini adalah untuk menentukan algoritma klasifikasi (K-NN atau Naive Bayes) yang secara performa komputasi

dan akurasi terbukti paling unggul dalam konteks studi kasus klasifikasi biji kopi.

1.4 Kerangka Kerja Penelitian (*Research Framework*)



Gambar 1.1 *Framework* Penelitian