

Penerapan Metode Naïve Bayes Pada Klasifikasi Gambar Digital Tumbuhan Obat Sirih Hijau dan Sirih Merah

Rizky Prabowo¹, Windy Desty Ariany² dan Bambang Hermanto³
Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Lampung
Jl. Prof. Dr. Ir. Sumantri Brojonegoro No.1 Kota Bandar Lampung, Lampung,
Indonesia
rizky.prabowo@fmipa.unila.ac.id, windy.desty1039@students.unila.ac.id,
bambang.hermanto@fmipa.unila.ac.id

Abstract

Plants possess numerous benefits as living organisms. For example, medicinal plants can be distinguished based on factors like their color, texture, and leaf structure. Categorization involves sorting similar plants together while noting their differences. Different techniques, such as the naive Bayes method, can be utilized for this purpose. This study evaluates the efficacy of naive Bayes classification in contrast to the Decision Tree approach. The dataset consists of original data extracted from 2000 images, which are divided into two groups: green betel and red betel. To facilitate comparison, the dataset is split into training and testing sets, with varying ratios ranging from 60:40 to 80:20.

Keywords: Medicinal Plants, Classification, Naïve Bayes.

Abstrak

Tanaman sebagai organisme memberikan sejumlah keunggulan yang signifikan. Misalnya, pengelompokan tanaman obat dapat dilakukan dengan menganalisis atribut-atribut seperti warna, tekstur, dan morfologi daun. Klasifikasi melibatkan pengelompokan tanaman dengan karakteristik serupa sambil membedakan yang berbeda. Berbagai metode dapat digunakan untuk klasifikasi, seperti metode naive Bayes. Dalam penelitian ini, kami mengevaluasi efektivitas klasifikasi naive Bayes dibandingkan dengan metode Pohon Keputusan. Dataset terdiri dari data asli yang diperoleh dari 2000 gambar, yang dikategorikan menjadi dua kelas: sirih hijau dan sirih merah. Dataset tersebut kemudian dibagi menjadi set data pelatihan dan pengujian, dengan berbagai rasio seperti 60:40, 65:35, 70:30, 75:25, dan 80:20 untuk tujuan perbandingan.

Kata kunci: Tumbuhan obat, Klasifikasi, Naïve Bayes.

1. PENDAHULUAN

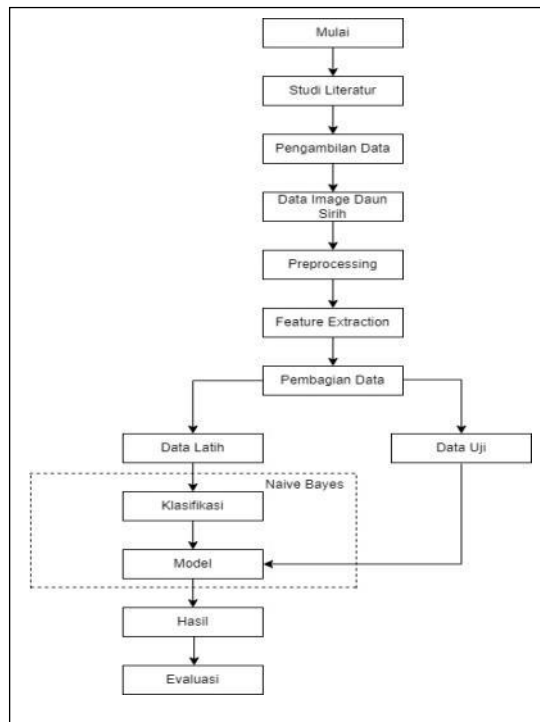
Tumbuhan menawarkan berbagai manfaat signifikan dan terdapat dalam beragam jenis, termasuk tumbuhan obat yang berperan penting sebagai sumber bahan makanan, obat-obatan, kosmetik, serta dalam pencegahan dan pengobatan berbagai kondisi kesehatan. Untuk mengidentifikasi tumbuhan obat, salah satu metode yang digunakan adalah dengan memfotografi daun herbalnya. Tumbuhan obat dikategorikan dengan memeriksa karakteristik seperti warna, tekstur, dan morfologi daun. Teknik ini melibatkan penggunaan metode pengolahan citra digital untuk mengklasifikasikan herbal berdasarkan struktur daunnya. Berbagai metode pengolahan citra digital telah dikembangkan untuk mendukung analisis citra, memfasilitasi pekerjaan manusia dalam pengolahan, analisis, dan pemanfaatan citra untuk beragam aplikasi dan kebutuhan. [1].

Klasifikasi merujuk pada prosedur mengelompokkan, yang melibatkan pengumpulan entitas atau objek serupa sementara memisahkannya dari yang berbeda. Istilah ini seringkali digunakan untuk menjelaskan metode penyusunan data yang sistematis dan presisi. Klasifikasi membagi data berdasarkan karakteristik spesifik. Grup dengan struktur yang mirip akan dikelompokkan bersama, menunjukkan kesesuaian dalam kategori yang sama. Proses klasifikasi dapat diterapkan melalui berbagai teknik, di antaranya adalah metode Naive Bayes. [2].

Metode klasifikasi *Naive Bayes* adalah pengklasifikasian dengan metode probabilitas dan statistik berdasarkan Teorema Bayes yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya[3]. Klasifikasi *Naive Bayes* dapat digunakan untuk memprediksi probabilitas keanggotaan dari suatu kelas. Ketika diterapkan pada *database* dalam jumlah besar, klasifikasi *Naive Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi[4].

2. METODOLOGI PENELITIAN

Dalam penelitian ini dilakukan dengan beberapa tahapan, yaitu dimulai dari studi literatur, proses pengambilan data, preprocessing, feature extraction, pembagian data, pemodelan dengan metode Naive Bayes hingga mendapatkan hasil, dan yang terakhir evaluasi, tahapan penelitian dapat dilihat pada gambar dibawah ini:



Gambar 1. Tahapan Penelitian

2.1 Studi Literatur

Peneliti menghimpun beragam sumber literatur, termasuk buku dan jurnal, yang relevan dengan objektif dan permasalahan studi mereka. Tujuan dari langkah ini adalah untuk menyediakan referensi yang akan memperkuat penelitian yang sedang dilakukan, berdasarkan pada karya-karya yang dihasilkan pada waktu sebelumnya.

2.2 Pengambilan Data

Data pada penelitian ini merupakan data primer yang diperoleh dari pengambilan gambar secara langsung. Data tersebut diperoleh dari 42 sample daun sirih hijau dan 42 sample daun sirih merah, sehingga total sample yang digunakan yaitu sebanyak 84 sample daun. Proses pengambilan gambar menggunakan kamera handphone, proses pengambilan gambar yaitu pada sample daun ke 1 hingga 41 difoto sebanyak 24 kali searah jarum jam, dan sample ke 42 difoto sebanyak 16 kali. Sehingga data yang dihasilkan dari 84 sample daun sebanyak 1000 data citra sirih hijau dan 1000 data citra sirih merah, sehingga seluruh data yang digunakan berjumlah 2.000 data citra.



Figure 2. Sirih Hijau



Figure 3. Sirih Merah

2.1 Preprocessing

Preprocessing merupakan langkah awal dalam pengolahan citra yang bertujuan untuk memperoleh citra yang optimal sebelum dilanjutkan ke tahap berikutnya. Proses preprocessing ini meliputi langkah resize, di mana gambar diubah ukurannya agar tetap memiliki rentang nilai yang sama, serta pemberian label pada data.

2.2 Feature Extraction

Ekstraksi fitur adalah proses penting dalam pengolahan data di berbagai bidang seperti pengolahan citra, pengenalan pola, dan pemrosesan bahasa alami. Tujuan utamanya adalah untuk mengubah data mentah menjadi representasi yang lebih sederhana dan informatif yang dapat digunakan untuk analisis lebih lanjut. Dalam konteks pengolahan citra, misalnya, ekstraksi fitur melibatkan identifikasi dan ekstraksi atribut-atribut penting dari gambar, seperti tepi, tekstur, warna, atau bentuk objek. Proses ekstraksi fitur dapat dilakukan dengan berbagai metode, termasuk metode statistik, transformasi domain spasial atau frekuensi, dan pendekatan berbasis model. Pemilihan metode yang tepat tergantung pada tugas yang akan diselesaikan dan karakteristik data yang dimiliki.

2.3 Pembagian Data

Data dipartisi menjadi data pelatihan dan data pengujian. Data pelatihan merujuk pada data yang digunakan untuk proses pembelajaran dari citra sirih hijau dan sirih merah, sementara data pengujian merujuk pada data yang digunakan untuk menguji hasil pembelajaran yang disimpan dalam model, berdasarkan citra sirih hijau dan sirih merah. Data tersebut merupakan gambar-gambar yang dipilih secara acak dari dataset citra yang digunakan selama proses pelatihan.

2.4 Pemodelan Naïve Bayes

Pada tahapan selanjutnya dalam penelitian ini yaitu proses klasifikasi menggunakan metode *Naïve Bayes*, penggunaan metode *Naïve Bayes* pada penelitian ini dikarenakan banyaknya dataset yang dipakai. Oleh karena itu dibutuhkan suatu metode yang memiliki kinerja klasifikasi kecepatan tinggi dan akurasi tinggi [5].

2.5 Hasil

Pada fase ini, dilakukan verifikasi terhadap hasil evaluasi akurasi klasifikasi citra sirih hijau dan sirih merah dengan menggunakan metode Naive Bayes untuk menilai tingkat ketepatannya. Hasil akurasi dari langkah ini memberikan indikasi apakah alat pemodelan yang digunakan dalam proses klasifikasi beroperasi dengan efektif atau tidak. Selain itu, tahap ini juga mencoba mengartikan hasil-hasil yang diperoleh dengan menganalisis keseluruhan data untuk mendapatkan kesimpulan terhadap semua objek yang dipelajari dengan menggunakan metode Naïve Bayes.

2.6 Evaluasi

Pada tahap ini, dilakukan evaluasi untuk mengukur seberapa tepat hasil klasifikasi citra tanaman obat menggunakan metode Naïve Bayes. Evaluasi dalam penelitian ini memanfaatkan Confusion Matrix sebagai sarana evaluasi, yang berguna untuk mengevaluasi kinerja klasifikasi serta mengukur akurasi pemodelan klasifikasi melalui parameter seperti akurasi, recall, presisi, dan skor F1. Hasil evaluasi ini akan mengindikasikan apakah alat pemodelan yang digunakan dalam proses klasifikasi beroperasi dengan baik atau tidak. Langkah ini juga bertujuan untuk menginterpretasikan hasil analisis dari seluruh data, dengan harapan dapat mengambil kesimpulan dari semua objek yang dianalisis menggunakan metode Naïve Bayes.

3. HASIL DAN PEMBAHASAN

3.1 Pemodelan 1

Dalam *experiment* awal pemodelan, data yang diambil memiliki ukuran 600x800 piksel, dan eksperimen dilakukan dengan membagi data menjadi 5 split. Persentase pembagian antara data pelatihan dan pengujian adalah 60:40, 65:35, 70:30, 75:25, dan 80:20. Pengujian model klasifikasi dengan *confusion matrix* dapat dilihat pada tabel 1.

Table 1. Pengujian Model Klasifikasi Pada Pemodelan 1

Experiment	Data	True Negatif	False Negative	True Positive	False Positive	Support
1	60 % latih 40 % uji	348	12	385	55	800
2	65 % latih 35 % uji	298	14	335	53	700
3	70 % latih 30 % uji	281	5	288	26	600
4	75 % latih 25 % uji	217	8	229	46	500
5	80 % latih 20 % uji	168	12	194	26	400

Tabel 1 merupakan hasil dari pengujian model klasifikasi menggunakan data berdimensi 600x800 *pixel*, dengan melakukan 5 kali *experiment*.

- Pada *experiment* pertama menggunakan 1200 data *train* dan 800 data *test*. Data *train* merupakan data yang digunakan untuk menjalankan fungsi dari algoritma. Sedangkan data *test* merupakan data uji yang digunakan untuk menguji performa dari algoritma. Dari pengujian tersebut didapatkan hasil *True Negative* = 348, *False Negative* = 12, *True Positive* = 385, *False Positive* = 55. Artinya dari *experiment* menggunakan 800 data uji, sebanyak 733 data terklasifikasi secara benar, dan 67 data terklasifikasikan secara salah.
- Pada *experiment* kedua menggunakan 1300 data *train* dan 700 data *test*. Data *train* merupakan data yang digunakan untuk menjalankan fungsi dari algoritma. Sedangkan data *test* merupakan data uji yang digunakan untuk menguji performa dari algoritma. Dari pengujian tersebut didapatkan hasil *True Negative* = 298, *False Negative* = 14, *True Positive* = 335, *False Positive* = 53. Artinya dari *experiment* menggunakan 700

data uji, sebanyak 633 data terklasifikasi secara benar, dan 67 data terklasifikasikan secara salah.

- Pada *experiment* ketiga menggunakan 1400 data *train* dan 600 data *test*. Data *train* merupakan data yang digunakan untuk menjalankan fungsi dari algoritma. Sedangkan data *test* merupakan data uji yang digunakan untuk menguji performa dari algoritma. Dari pengujian tersebut didapatkan hasil *True Negative* = 281, *False Negative* = 5, *True Positive* = 288, *False Positive* = 26. Artinya dari *experiment* menggunakan 600 data uji, sebanyak 569 data terklasifikasi secara benar, dan 31 data terklasifikasikan secara salah.
- Pada *experiment* keempat menggunakan 1500 data *train* dan 500 data *test*. Data *train* merupakan data yang digunakan untuk menjalankan fungsi dari algoritma. Sedangkan data *test* merupakan data uji yang digunakan untuk menguji performa dari algoritma. Dari pengujian tersebut didapatkan hasil *True Negative* = 217, *False Negative* = 9, *True Positive*= 229, *False Positive* = 46. Artinya dari *experiment* menggunakan 500 data uji, sebanyak 446 data terklasifikasi secara benar, dan 54 data terklasifikasikan secara salah.
- Pada *experiment* kelima menggunakan 1600 data *train* dan 400 data *test*. Data *train* merupakan data yang digunakan untuk menjalankan fungsi dari algoritma. Sedangkan data *test* merupakan data uji yang digunakan untuk menguji performa dari algoritma. Dari pengujian tersebut didapatkan hasil *True Negative*= 168, *False Negative* = 12, *True Positive*= 194, *False Positive* = 26. Artinya dari *experiment* menggunakan 400 data uji, sebanyak 362 data terklasifikasi secara benar, dan 38 data terklasifikasikan secara salah.

Setelah mendapat hasil dari *confusion matrix* maka didapatkalah hasil evaluasi kinerja klasifikasi, hasil evaluasi dapat dilihat pada tabel 2.

Table 2. Evaluasi Kinerja Klasifikasi Pada Pemodelan 1

<i>Experiment</i>	Data	<i>Accu</i>	<i>Prec</i>	<i>Rec</i>	<i>F1</i>
1	60 % <i>train</i> 40 % <i>test</i>	91,63%	87,50%	96,98%	92,00%
2	65 % <i>train</i> 35 % <i>test</i>	90,43%	86,34%	95,99%	90,91%
3	70 % <i>train</i> 30 % <i>test</i>	94,83%	91,72%	98,29%	94,89%

4	75 % <i>train</i> 25 % <i>test</i>	89,20%	83,27%	96,62%	89,45%
5	80 % <i>train</i> 20 % <i>test</i>	90,50%	88,18%	94,17%	91,08%

Tabel 2 merupakan tabel evaluasi kinerja klasifikasi pada pemodelan yang dilakukan. Berikut hasil evaluasi kinerja klasifikasi yang didapatkan dari ke 5 *experiment* tersebut.

- Pada *experiment* pertama evaluasi kinerja klasifikasi didapatkan *Accuracy* 91,63%, *Precision* 87,50%, *Recall* 96,98% dan *F1 Score* 92,00%.
- Pada *experiment* kedua evaluasi kinerja klasifikasi didapatkan *Accuracy* 90,43%, *Precision* 86,34%, *Recall* 95,99% dan *F1 Score* 90,91%.
- Pada *experiment* ketiga evaluasi kinerja klasifikasi didapatkan *Accuracy* 94,83%, *Precision* 91,72%, *Recall* 98,29% dan *F1 Score* 94,89%
- Pada *experiment* keempat evaluasi kinerja klasifikasi didapatkan *Accuracy* 89,20%, *Precision* 83,27%, *Recall* 96,62% dan *F1 Score* 89,45%
- Pada *experiment* kelima evaluasi kinerja klasifikasi didapatkan *Accuracy* 90,50%, *Precision* 88,18%, *Recall* 94,17% dan *F1 Score* 91,08%

4. SIMPULAN

Pada penelitian ini metode *Naïve Bayes* sudah berhasil diimplementasikan untuk mengklasifikasikan tumbuhan obat sirih hijau dan sirih merah. Hasil klasifikasi dari metode *Naïve Bayes* dilakukan dengan menggunakan dataset berdimensi 600x800 pixel yang dilakukan dengan 5 kali *experiment*. Dari hasil rata-rata didapatkan *Accuracy Naïve Bayes* sebesar 91,32%, *Precision Naïve Bayes* sebesar 87,40%, *Recall Naïve Bayes* sebesar 96,41%, nilai *F1-Score Naïve Bayes* sebesar 91,67%. Jika dilihat secara umum *naïve bayes* dapat mengklasifikasikan sirih merah dan sirih hijau dengan baik. Penilaian secara keseluruhan rata-rata di atas 85%.

DAFTAR PUSTAKA

- [1] Liantoni, F., & Nugroho, H. (2015). **Klasifikasi Daun Herbal Menggunakan Metode Naïve Bayes Classifier Dan Knearest Neighbor.** *Jurnal Simantec*, 5(1), 9-16.
- [2] D Kurniawan, A Aristoteles, A. A. (2017). **Pengembangan Aplikasi Sistem Pembelajaran Klasifikasi (Taksonomi) dan Tata Nama Ilmiah (Binomial Nomenklatur) pada Kingdom Plantae (Tumbuhan) Berbasis Android.** *Journal of Chemical Information and Modeling*, 4(2), 1-15. <https://jurnal.fmipa.unila.ac.id/komputasi/article/view1143/937>
- [3] Saleh, A., & Utama, U. P. (2016). **Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga.** *January 2015*.
- [4] Miranda, E., & Julisar. (2018). **Data mining dengan metode klasifikasi naïve bayes untuk mengklasifikasikan pelanggan** Eka Miranda , Julisar Program Sistem Informasi , Program Studi Sistem Informasi , Universitas Bina Nusantara. *Infotech*, 4(9), 6-12.
- [5] Nurhuda, F., Widya Sihwi, S., & Doewes, A. (2016). **Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier.** *Jurnal Teknologi & Informasi ITSmart*, 2(2), 35. <https://doi.org/10.20961/its.v2i2.630>