Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer
Volume 4, Nomor 2, Juli 2025
p-ISSN: 2827-8135; e-ISSN: 2827-7953, Hal. 207-213
DOI: https://doi.org/10.55606/juisik.v4i2.1228
Available Online at: https://journal.sinov.id/index.php/juisik

Klasifikasi Jenis Bunga Iris Menggunakan Algoritma Klasifikasi Tradisional

Alwi Syahputra [1*], Rusma Riansyah [2], Dimas Aqila Aptanta [3], Muhammad Farhan [4], Mhd. Furqan [5]
[1-5] Universitas Islam Negeri Sumatera Utara, Indonesia
alwiisyahputra0@gmail.com [1*], riansyahrusma@gmail.com [2], dimasaqila71@gmail.com [3], muhammadfarhan3737iyan@gmail.com [4], mfurqan@gmail.com [5]

Alamat: Jl Lap Golf, Kec Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara
Korespodensi email: alwiisyahputra0@gmail.com

Abstract. *This study aims to implement and compare the performance of two traditional classification algorithms, namely K-Nearest Neighbor (K-NN) and Naive Bayes to classify Iris flower types. The dataset used is the Iris Dataset which is a classic dataset in machine learning consisting of 150 samples with four features (sepal length, sepal width, petal length, and petal width) and three target classes (Iris Setosa, Iris Versicolor, and Iris Virginica). The research methodology includes data preprocessing, algorithm implementation, model evaluation using accuracy, precision, recall, and F1-score metrics, and comparative performance analysis. The results showed that the K-NN algorithm with k = 3 achieved an accuracy of 96.67%, while Naive Bayes achieved an accuracy of 93.33%. Both algorithms showed good performance in classifying Iris flower types, with K-NN slightly superior in terms of accuracy. This study proves that traditional classification algorithms are still relevant and effective for classification problems with less complex datasets.*

Keywords: *Iris Dataset, Klasifikasi, K-Nearest Neighbor, Machine Learning, Naïve Bayes*

Abstrak. Penelitian ini bertujuan untuk mengimplementasikan dan membandingkan performa dua algoritma klasifikasi tradisional yaitu K-Nearest Neighbor (K-NN) dan Naive Bayes untuk mengklasifikasikan jenis bunga Iris. Dataset yang digunakan adalah Iris Dataset yang merupakan dataset klasik dalam machine learning yang terdiri dari 150 sampel dengan empat fitur (panjang sepal, lebar sepal, panjang petal, dan lebar petal) dan tiga kelas target (Iris Setosa, Iris Versicolor, dan Iris Virginica). Metodologi penelitian meliputi preprocessing data, implementasi algoritma, evaluasi model menggunakan metrik akurasi, precision, recall, dan F1-score, serta analisis perbandingan performa. Hasil penelitian menunjukkan bahwa algoritma K-NN dengan k=3 mencapai akurasi 96.67%, sedangkan Naive Bayes mencapai akurasi 93.33%. Kedua algoritma menunjukkan performa yang baik dalam mengklasifikasikan jenis bunga Iris, dengan K-NN sedikit lebih unggul dalam hal akurasi. Penelitian ini membuktikan bahwa algoritma klasifikasi tradisional masih relevan dan efektif untuk masalah klasifikasi dengan dataset yang tidak terlalu kompleks.

Kata kunci: Iris Dataset, Klasifikasi, K-Nearest Neighbor, Machine Learning, Naïve Bayes

1. LATAR BELAKANG

Klasifikasi merupakan salah satu tugas fundamental dalam machine learning yang bertujuan untuk memprediksi kelas atau kategori dari suatu data berdasarkan fitur-fitur yang dimilikinya (Saputro et al., 2024). Dalam bidang botani, klasifikasi otomatis jenis tanaman atau bunga dapat membantu para peneliti dan praktisi dalam mengidentifikasi spesies dengan lebih cepat dan akurat (Annurullah et al., 2024).

Dataset Iris yang diperkenalkan oleh Ronald Fisher pada tahun 1936 telah menjadi benchmark klasik dalam bidang machine learning dan pattern recognition (Wulandari et al., 2024). Dataset ini berisi pengukuran morfologi bunga Iris dari tiga spesies berbeda: Iris

Received: Mei 12, 2025; Revised: Mei 30, 2025; Accepted: Juni 17, 2025; Online Available: Juni 19, 2025

Setosa, Iris Versicolor, dan Iris Virginica. Setiap sampel memiliki empat fitur numerik yang merepresentasikan dimensi fisik bunga (Hidayat & Sari, 2024).

Algoritma klasifikasi tradisional seperti K-Nearest Neighbor (K-NN) dan Naive Bayes telah terbukti efektif dalam berbagai aplikasi klasifikasi (Rahman & Putri, 2024; Susanto & Wijaya, 2023). K-NN merupakan algoritma lazy learning yang mengklasifikasikan data berdasarkan kedekatan dengan data training, sedangkan Naive Bayes menggunakan pendekatan probabilistik berdasarkan teorema Bayes dengan asumsi independensi antar fitur (Arifin & Ariesta, 2023).

Penelitian ini bertujuan untuk mengimplementasikan dan membandingkan performa kedua algoritma tersebut dalam mengklasifikasikan jenis bunga Iris, serta memberikan analisis mendalam tentang kelebihan dan kekurangan masing-masing metode.

2. KAJIAN TEORITIS

K-Nearest Neighbor (K-NN)

K-Nearest Neighbor adalah algoritma supervised learning yang sederhana namun powerful untuk masalah klasifikasi dan regresi (Pratama & Sari, 2023). Algoritma ini bekerja berdasarkan prinsip bahwa data yang mirip cenderung memiliki kelas yang sama. Proses klasifikasi dilakukan dengan mencari k tetangga terdekat dari data uji dalam ruang fitur, kemudian menentukan kelas berdasarkan voting mayoritas dari k tetangga tersebut (Simbolon et al., 2022).

Beberapa penelitian terkini menunjukkan efektivitas K-NN dalam berbagai domain (Fernanda & Fathoni, 2022). Algoritma ini memiliki beberapa keunggulan seperti kesederhanaan implementasi, tidak memerlukan asumsi distribusi data, dan dapat menangani data multi-kelas dengan baik. Namun, K-NN juga memiliki kelemahan seperti computational cost yang tinggi pada dataset besar dan sensitivitas terhadap curse of dimensionality (Annurullah et al., 2024).

Naive Bayes

Naive Bayes adalah algoritma klasifikasi probabilistik yang berdasarkan pada teorema Bayes dengan asumsi "naive" bahwa fitur-fitur dalam dataset bersifat independen satu sama lain (Hikmah et al., 2024). Meskipun asumsi independensi ini jarang terpenuhi dalam praktik, Naive Bayes sering menunjukkan performa yang mengejutkan dalam berbagai aplikasi (Rahman & Putri, 2024).

Algoritma ini memiliki beberapa varian seperti Gaussian Naive Bayes untuk fitur kontinu, Multinomial Naive Bayes untuk fitur diskrit, dan Bernoulli Naive Bayes untuk fitur

biner (Arifin & Ariesta, 2023). Keunggulan Naive Bayes meliputi kecepatan training dan prediksi yang tinggi, kemampuan menangani missing values, dan performa yang baik pada dataset kecil (Pratama & Sari, 2023).

Dataset Iris

Dataset Iris merupakan salah satu dataset paling terkenal dalam komunitas machine learning (Arhami & Nasir, 2020). Dataset ini terdiri dari 150 sampel yang terbagi rata menjadi tiga kelas (50 sampel per kelas). Empat fitur yang terdapat dalam dataset adalah panjang sepal, lebar sepal, panjang petal, dan lebar petal, semuanya dalam satuan sentimeter (Dinata & Hasdyna, 2020).

Karakteristik unik dari dataset Iris adalah adanya separabilitas linear yang baik antara kelas Iris Setosa dengan dua kelas lainnya, sementara Iris Versicolor dan Iris Virginica memiliki overlap yang signifikan (Simbolon et al., 2022). Hal ini menjadikan dataset Iris sebagai benchmark yang ideal untuk menguji algoritma klasifikasi (Santoso et al., 2021).

3. METODE PENELITIAN

Dataset dan Preprocessing

Dataset Iris yang digunakan dalam penelitian ini diakses melalui library scikit-learn Python (Raharjo & Purnama, 2024). Dataset terdiri dari 150 sampel dengan distribusi yang seimbang antar kelas. Tahap preprocessing meliputi:

- Eksplorasi Data: Analisis statistik deskriptif, visualisasi distribusi fitur, dan identifikasi outlier (Hikmah et al., 2024).
- Normalisasi: Standardisasi fitur menggunakan StandardScaler untuk memastikan semua fitur memiliki skala yang sama (Arhami & Nasir, 2020).
- Pembagian Data: Split dataset menjadi 80% data training (120 sampel) dan 20% data testing (30 sampel) menggunakan stratified sampling (Dinata & Hasdyna, 2020).

Implementasi Algoritma

- 3.2.1 K-Nearest Neighbor

Implementasi K-NN menggunakan library scikit-learn dengan parameter sebagai berikut (Santoso et al., 2021):

- Jumlah tetangga (k): Dilakukan eksperimen dengan k = 1, 3, 5, 7, 9
- Fungsi jarak: Euclidean distance
- Pembobotan: Uniform weighting
- 3.2.2 Naive Bayes

Implementasi Naive Bayes menggunakan Gaussian Naive Bayes yang sesuai untuk fitur kontinu (Arifin & Ariesta, 2023). Parameter yang digunakan:

- Prior probability: Uniform (seimbang antar kelas)
- Variance smoothing: 1e-9 (default)

Evaluasi Model

Evaluasi dilakukan menggunakan beberapa metrik (Pratama & Sari, 2023):

- Akurasi: Rasio prediksi benar terhadap total prediksi
- Precision: Rasio true positive terhadap predicted positive
- Recall: Rasio true positive terhadap actual positive
- F1-Score: Harmonic mean dari precision dan recall
- Confusion Matrix: Visualisasi performa klasifikasi per kelas

Cross-Validation: 10-fold cross-validation untuk validasi robustness (Fernanda & Fathoni, 2022).

4. HASIL DAN PEMBAHASAN

Analisis Eksploratif Data

Analisis statistik deskriptif menunjukkan bahwa fitur petal length dan petal width memiliki variabilitas yang lebih tinggi dibandingkan fitur sepal (Saputro et al., 2024). Visualisasi scatter plot menunjukkan separabilitas yang jelas antara Iris Setosa dengan dua kelas lainnya, sementara Iris Versicolor dan Iris Virginica menunjukkan overlap terutama pada region boundary (Simbolon et al., 2022).

Analisis korelasi menunjukkan korelasi positif yang kuat antara petal length dan petal width (r=0.96), serta korelasi moderat antara sepal length dan petal length (r=0.87) (Hikmah et al., 2024). Hal ini mengindikasikan adanya redundansi informasi yang dapat dimanfaatkan oleh algoritma klasifikasi.

Performa K-Nearest Neighbor

Eksperimen dengan berbagai nilai k menunjukkan hasil sebagai berikut:

Tabel 1. Hasil Performa K-Nearest Neighbor

K	Akurasi (%)	Precision (%)	Recall (%)	F1-Score (%)
1	93.33	93.67	93.33	93.44
3	96.67	96.89	96.67	96.74
5	96.67	96.89	96.67	96.74
7	93.33	93.67	93.33	93.44
9	90.00	90.48	90.00	90.16

Nilai k=3 dan k=5 memberikan performa terbaik dengan akurasi 96.67% (Annurullah et al., 2024). Analisis confusion matrix menunjukkan bahwa semua sampel Iris Setosa diklasifikasikan dengan benar, sementara terdapat satu misklasifikasi antara Iris Versicolor dan Iris Virginica (Fernanda & Fahoni, 2022).

Performa Naive Bayes

Gaussian Naive Bayes menunjukkan performa yang konsisten dengan hasil sebagai berikut:

Tabel 2. Hasil Performa Naive Bayes

Metrik	Nilai (%)
Akurasi	93.33
Precision	94.44
Recall	93.33
F1-Score	93.60

Analisis per kelas menunjukkan bahwa Iris Setosa diklasifikasikan dengan sempurna (100% akurasi), sementara Iris Versicolor dan Iris Virginica memiliki beberapa kesalahan klasifikasi (Rahman & Putri, 2024). Hal ini konsisten dengan karakteristik dataset yang menunjukkan overlap antara kedua kelas tersebut (Simbolon et al., 2022).

Perbandingan Algoritma

Perbandingan komprehensif antara kedua algoritma:

Tabel 3. Hasil Perbandingan Komprehensif

Aspek	K-NN (k=3)	Naive Bayes
Akurasi	96.67%	93.33%
Waktu Training	0.001s	0.002s
Waktu Prediksi	0.003s	0.001s
Memory Usage	Medium	Low
Interpretabilitas	Low	High
Robustness	Medium	High

Cross-Validation Results

Hasil 10-fold cross-validation menunjukkan (Susanto & Wijaya, 2023):

- K-NN (k=3): Mean accuracy = 95.33% ± 4.22%
- Naive Bayes: Mean accuracy = 94.00% ± 5.16%

Kedua algoritma menunjukkan konsistensi yang baik dengan standard deviation yang relatif rendah (Pratama & Sari, 2023).

Analisis Feature Importance

Analisis kontribusi fitur menggunakan permutation importance menunjukkan bahwa petal length dan petal width merupakan fitur yang paling diskriminatif, diikuti oleh sepal length dan sepal width (Wulandari et al., 2024). Hal ini konsisten dengan domain knowledge bahwa karakteristik petal lebih bervariasi antar spesies Iris (Arhami & Nasir).

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan serta membandingkan dua algoritma klasifikasi tradisional, yaitu K-Nearest Neighbors (K-NN) dan Naive Bayes, dalam klasifikasi jenis bunga Iris. Hasilnya menunjukkan bahwa K-NN dengan parameter *k=3* memberikan akurasi tertinggi sebesar 96,67%, sedikit lebih unggul dibandingkan Naive Bayes yang mencapai 93,33%. Meskipun perbedaannya relatif kecil, kedua algoritma menunjukkan performa yang sebanding. Dari segi karakteristik, K-NN lebih sensitif terhadap pemilihan parameter dan membutuhkan komputasi lebih tinggi pada tahap prediksi, sedangkan Naive Bayes lebih cepat dan mudah diinterpretasikan. Keduanya mampu mengklasifikasikan Iris Setosa secara sempurna, mengindikasikan separabilitas linear yang tinggi, meskipun masih terdapat kesalahan pada batas klasifikasi antara Iris Versicolor dan Iris Virginica. Uji *cross-validation* juga menunjukkan bahwa keduanya memiliki stabilitas prediksi yang baik dengan variasi rendah. Secara praktis, untuk dataset dengan ukuran kecil, fitur numerik, dan distribusi kelas yang seimbang, kedua algoritma ini sangat layak digunakan karena kesederhanaan implementasi dan efisiensi prediksinya. Studi ini menegaskan bahwa algoritma klasifikasi tradisional tetap relevan untuk permasalahan klasifikasi sederhana, dan merekomendasikan eksplorasi lebih lanjut terhadap metode *ensemble*, teknik seleksi fitur, serta aplikasi pada dataset yang lebih besar dan kompleks di masa mendatang.

DAFTAR REFERENSI

Annurullah, G. A., Mulya, A., Dermawan, T., Haiban, R. R., Tatamara, A., & Khalifah, H. D. (2024). Analisis perbandingan KNN, SVM, Decision Tree dan Regresi Logistik untuk klasifikasi obesitas multi kelas. *KLIK: Kajian Ilmiah Informatika dan Komputer, 5*(1), 102–115.

Arhami, M., & Nasir, M. (2020). *Data mining: Algoritma dan implementasi*. Yogyakarta: Penerbit Andi.

Arifin, T., & Ariesta, D. (2023). Prediksi penyakit ginjal kronis menggunakan algoritma Naive Bayes classifier berbasis particle swarm optimization. *Jurnal Tekno Insentif, 17*(1), 26–35.

Dinata, R. K., & Hasdyna, N. (2020). *Machine learning*. Lhokseumawe: Unimal Press.

Fernanda, M., & Fathoni, N. (2022). Perbandingan klasifikasi antara KNN dan Naive Bayes pada penentuan status gunung berapi dengan K-fold cross validation. *Jurnal Informatika dan Teknologi, 15*(3), 45–58.

Hidayat, R., & Sari, M. (2024). Analisis perbandingan algoritma Naive Bayes dan KNN untuk

analisis sentimen ulasan pengguna aplikasi Vidio di Google Play Store. *CONTEN: Computer and Network Technology, 4*(2), 78–89.

Hikmah, A. B., Susanto, R., & Wahyudi, M. (2024). *Algoritma data science*. Jakarta: Teknosain.

Pratama, Y., & Sari, D. P. (2023). Analisis performa metode klasifikasi Naïve Bayes classifier pada unbalanced dataset. *Indonesian Journal of Data and Science, 4*(3), 89–102.

Raharjo, S., & Purnama, I. K. E. (2024). *Buku ajar data mining*. Surabaya: Airlangga University Press.

Rahman, A., & Putri, S. A. (2024). Klasifikasi kategori berita menggunakan Naive Bayes. In *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)* (Vol. 7, No. 1, pp. 234–242).

Santoso, B., Wijaya, A., & Kurniawan, D. (2021). *Algoritma machine learning*. Jakarta: Elex Media Komputindo.

Saputro, W. E., Yuana, H., & Puspitasari, W. D. (2024). Perbandingan metode K-Nearest Neighbor (KNN) dan Naive Bayes terhadap analisis sentimen pada pengguna e-wallet aplikasi DANA menggunakan fitur ekstraksi TF-IDF. *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika, 18*(2), 45–58.

Simbolon, H., Turnip, M., & Sitompul, O. S. (2022). The klasifikasi kualitas air menggunakan metode KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi), 9*(2), 1210–1223.

Susanto, B., & Wijaya, C. (2023). Implementasi algoritma Naïve Bayes classifier dan K-nearest neighbor untuk klasifikasi penyakit ginjal kronik. *MALCOM: Indonesian Journal of Machine Learning and Computer Science, 3*(2), 156–167.

Wulandari, A., Mulya, A., Dermawan, T., Haiban, R. R., Tatamara, A., & Khalifah, H. D. (2024). Application of artificial neural network, K-nearest neighbor and Naive Bayes algorithms for classification of obesity risk cardiovascular disease. *IJATIS: Indonesian Journal of Applied Technology and Innovation Science, 1*(1), 9–15.