

# Analisis Klasifikasi Spesies Bunga Iris Menggunakan Algoritma K-Nearest Neighbors

<sup>1\*</sup>Moh. Hasan Efendi, <sup>2</sup>Wildan Septian Pratama, <sup>3</sup>Erna Daniati

<sup>1-3</sup> Sistem Informasi, Universitas Nusantara PGRI Kediri

E-mail: <sup>1</sup>[hasanefendi1258@gmail.com](mailto:hasanefendi1258@gmail.com), <sup>2</sup>[wsseptama@gmail.com](mailto:wseptama@gmail.com), <sup>3</sup>[ernadaniati@unpkediri.ac.id](mailto:ernadaniati@unpkediri.ac.id)

*Penulis Korespondens : Erna Daniati*

**Abstrak**—Penelitian ini bertujuan untuk mengklasifikasikan spesies bunga Iris menggunakan pendekatan K-Nearest Neighbors (KNN). Topik ini menarik karena adanya penerapan teknik mesin learning dalam botani, serta juga karena dataset Iris adalah dataset yang memiliki struktur jelas dan seringkali digunakan sebagai benchmark dalam penelitian. Dalam pembuatan model ini, saya melakukan eksplorasi data dan visualisasi untuk mendapatkan understanding, pemisahan data dengan teknik pemisahan data train-test, dan pelatihan serta evaluasi model KNN dengan beragam nilai k. Berdasarkan analisis yang telah kami lakukan, pada model ini pembelajaran dengan k=11 memperoleh akurasi 96,67% pada klasifikasi tiga jenis spesies bunga Iris. Evaluasi model dengan akurasi, precision, recall, dan F1-score menunjukkan hasil yang konsisten sangat baik. Temuan ini menegaskan bahwa algoritma KNN diimplementasikan pada dataset sederhana secara efektif digunakan untuk klasifikasi, dan masih tetap bisa digunakan pada tipe data lainnya untuk aplikasi lain.

**Kata Kunci**— bunga iris, klasifikasi, K-Nearest Neighbors, machine learning

**Abstract**—This research aims to classify Iris flower species using the K-Nearest Neighbors (KNN) approach. This topic is interesting due to the application of machine learning techniques in botany, as well as because the Iris dataset is a well-structured dataset that is frequently used as a benchmark in research. In building this model, I performed data exploration and visualization to gain an understanding, data splitting using the train-test split technique, and training and evaluating the KNN model with various values of k. Based on the analysis we conducted, the model with k=11 achieved an accuracy of 96.67% in classifying the three Iris flower species. Model evaluation, including accuracy, precision, recall, and F1-score, showed consistently excellent results. These findings confirm that the KNN algorithm, when implemented on a simple dataset, can be effectively used for classification and can still be applied to other types of data for different applications.

**Keywords**— iris flower, classification, K-Nearest Neighbors, machine learning

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



## I. PENDAHULUAN

Klasifikasi spesies bunga Iris (Iris spp.) berdasarkan karakteristik morfologis telah menjadi studi kasus populer dalam bidang machine learning karena kompleksitas pola distribusi fitur antarspesies. Meskipun perbedaan visual antar Iris setosa, Iris versicolor, dan Iris virginica dapat dikenali ahli botani, identifikasi manual menjadi tidak efisien saat menghadapi dataset besar atau variasi fenotip yang subtil.[1] Tantangan utama terletak pada optimalisasi akurasi klasifikasi ketika fitur-fitur seperti panjang sepal, lebar petal, dan rasio kelopak menunjukkan tumpang tindih antarspesies, khususnya antara Iris versicolor dan Iris virginica.[2]

Beberapa penelitian terdahulu telah menguji efektivitas algoritma K-Nearest Neighbors (KNN) untuk kasus ini. Rahman et al. (2024) [2] berhasil mencapai akurasi 100% menggunakan KNN pada platform RapidMiner dengan dataset Fisher's Iris yang terstandarisasi. Namun, literatur terkini mengidentifikasi dua kesenjangan: (1) ketergantungan kinerja KNN pada teknik preprocessing seperti normalisasi Z-Score, dan (2) terbatasnya eksplorasi parameter optimal K dalam konteks dataset Iris.[3]

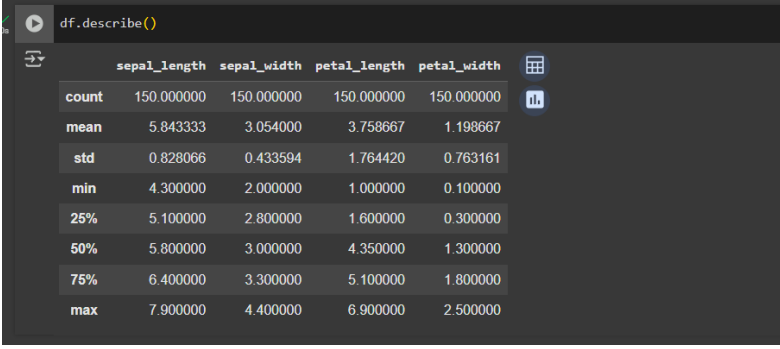
Penelitian ini bertujuan menganalisis sensitivitas algoritma KNN terhadap variasi parameter K dan strategi normalisasi data, sekaligus memvalidasi konsistensi performanya lintas platform komputasi. Hipotesis utama menyatakan bahwa kombinasi seleksi fitur morfologis kunci (panjang petal dan rasio sepal-petal) dengan optimasi nilai K akan menghasilkan akurasi klasifikasi melebihi 95% meskipun menggunakan alat analisis berbasis open-source.[4] Temuan ini diharapkan dapat memperkuat fondasi teoritis penerapan KNN untuk klasifikasi taksonomi tanaman sekaligus memberikan panduan praktis dalam pemilihan parameter algoritma.[5]

## II. METODE

### 2.1 Pemilihan Dataset

Penelitian ini menggunakan dataset Iris yang diolah dengan Python di Google Colab. Dataset ini berisi 150 data bunga dari tiga spesies: Iris Setosa, Virginica, dan Versicolor. Masing-masing sampel memiliki empat fitur utama, yaitu panjang dan lebar sepal serta petal (dalam sentimeter). Dataset ini tersedia secara bebas dan dapat diakses melalui Kaggle atau UCI Machine Learning Repository, dan digunakan untuk keperluan klasifikasi dalam penelitian ini.[6]

Sebelum memulai proses klasifikasi, langkah pertama yang dilakukan adalah eksplorasi dan visualisasi data untuk memahami distribusi serta karakteristik setiap spesies.[7] Pada tahap ini, data akan dianalisis menggunakan statistik deskriptif untuk memperoleh informasi mengenai nilai



	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

Gambar 1 : Dataset

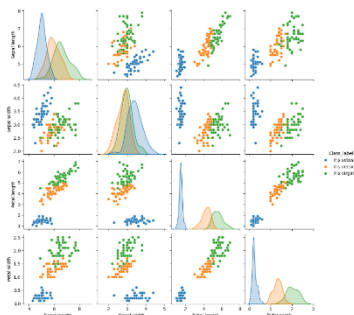
rata-rata, standar deviasi, dan tingkat korelasi antara fitur-fitur yang ada. Analisis ini dilakukan seperti yang dijelaskan berikut:

Untuk memvisualisasikan data, digunakan scatter plot dalam bentuk *pair plot*. [8] Visualisasi ini sangat membantu karena menampilkan hubungan antar semua fitur dalam satu gambar. Dari sini terlihat jelas bahwa Iris-setosa memiliki petal yang lebih kecil dan data yang lebih terkonsentrasi, sedangkan Iris-versicolor dan Iris-virginica memiliki sebaran yang lebih luas, terutama pada fitur petal.[9] Ini menunjukkan bahwa ukuran petal cukup efektif membedakan Iris-setosa dari dua spesies lainnya, meskipun belum cukup kuat untuk memisahkan versicolor dan virginica secara sempurna. Pair plot ini memberikan banyak insight penting dari data secara visual.[10]

Hubungan Antar Fitur: Pair plot menampilkan scatter plot untuk setiap pasangan fitur dalam dataset, yang memungkinkan kita untuk melihat bagaimana setiap fitur saling berhubungan. Ini memudahkan identifikasi korelasi antar fitur. Selain itu, pair plot juga menampilkan histogram atau KDE (Kernel Density Estimation) pada diagonal, yang menggambarkan distribusi dari masing-masing fitur. Hal ini membantu kita memahami bagaimana data tersebar untuk setiap fitur individual[11]

```
sns.pairplot(df, hue='Class_labels', diag_kind='kde')  
plt.show()
```

Gambar 2 : Pair Plot



Gambar 3 : Visualisasi data

## 2.2 Penerapan Klasifikasi Algoritma K-Nearest Neighbors (KNN)

Algoritma K-Nearest Neighbors (K-NN) adalah salah satu metode pembelajaran mesin berbasis instance yang digunakan untuk klasifikasi dan regresi. [12]

### Langkah-langkah Implementasi Klasifikasi K-NN:

#### 1. Mempersiapkan Data:

- Dataset yang digunakan adalah **Iris Dataset**, yang mencakup 4 fitur: panjang dan lebar sepal serta panjang dan lebar petal. Dataset ini memiliki tiga kelas bunga, yaitu *Iris-setosa*, *Iris-versicolor*, dan *Iris-virginica*.

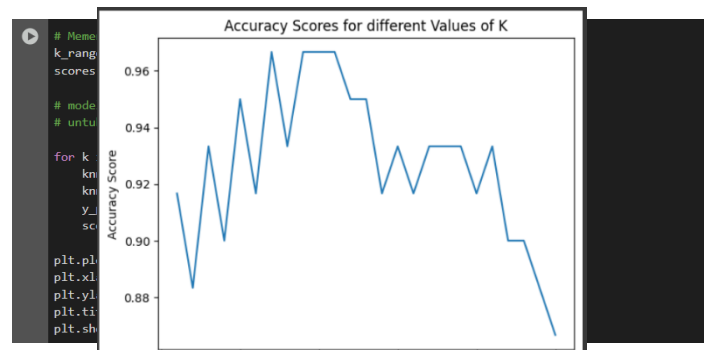
#### 2. Pembagian Data:

60% data dipilih secara acak untuk digunakan sebagai data pelatihan (training data), dan 40% sisanya digunakan sebagai data pengujian (testing data).[13] Pembagian ini bertujuan untuk mengevaluasi kinerja model pada data yang belum dilihat sebelumnya. Kode untuk pembagian data dapat dilihat pada gambar berikut:

```
[ ] # 60% data akan dipilih acak saat training data  
# remaining 40% sebagai testing data  
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=0)
```

Gambar 4 : Training Data

3. **Menentukan Nilai k:** Nilai k adalah jumlah tetangga terdekat yang digunakan untuk klasifikasi. Dalam percobaan ini, kami menguji berbagai nilai k dalam rentang 1 hingga 26 untuk melihat akurasi model pada setiap nilai k. Hasil akurasi tersebut diukur dan digambarkan dalam grafik berikut:



Gambar 5: Menentukan Nilai k

Gambar 6 : Grafik Akurasi Skor

Grafik menunjukkan variasi skor akurasi yang dihitung berdasarkan berbagai nilai k, dan memberikan gambaran tentang nilai k yang optimal untuk dataset ini.

#### 4. Menggunakan k-NN untuk Klasifikasi:

Setelah memilih nilai k yang optimal berdasarkan hasil grafik, k-NN diterapkan pada data pelatihan. Model dilatih menggunakan fitur  $x_{\text{train}}$  dan label  $y_{\text{train}}$ , kemudian diuji pada data pengujian ( $x_{\text{test}}$ ). [14]

Berikut adalah implementasi model k-NN dengan  $k=11$  (nilai optimal berdasarkan percobaan):

```
# Optimal K-Value
knn = KNeighborsClassifier(n_neighbors=11)
knn.fit(x_train, y_train)
print(knn.score(x_test, y_test))
```

0.9666666666666667

Gambar 7 : Hasil Akurasi

Dengan nilai  $k=11$ , model menghasilkan akurasi 96.67% pada data pengujian.

#### 3.4 Proses Pembagian Data

Proses pembagian data dilakukan dengan menggunakan metode *train-test split* untuk membagi dataset menjadi dua bagian: data pelatihan dan data pengujian. Pembagian ini dilakukan dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian, bertujuan agar model yang dikembangkan dapat diuji menggunakan data yang belum pernah ditemui sebelumnya [15]. Seluruh pembagian dilakukan secara acak dengan memperhatikan distribusi kelas yang seimbang.

#### 3.5 Pelatihan Model

Setelah data dibagi, langkah berikutnya adalah melatih model dengan menggunakan data pelatihan. Setiap algoritma yang telah dipilih akan diterapkan pada data pelatihan untuk membangun model klasifikasi. Proses ini mencakup penentuan parameter optimal untuk setiap

model melalui teknik validasi silang (cross-validation) guna memastikan bahwa hasil yang diperoleh lebih dapat dipercaya[16].

### 3.6. Evaluasi Kinerja Model

Kinerja tiap model dievaluasi menggunakan beberapa metrik, yaitu:

- **Akurasi:** persentase prediksi yang benar,
  - **Precision:** ketepatan prediksi positif,
  - **Recall:** sejauh mana model mendeteksi seluruh kasus positif,
  - **F1-score:** rata-rata harmonis antara precision dan recall.
- Evaluasi dilakukan menggunakan data pengujian yang tidak digunakan saat pelatihan, agar hasilnya lebih objektif.

## III. HASIL DAN PEMBAHASAN

### 3.11 Algoritma Klasifikasi – K-Nearest Neighbors (k-NN)

Berdasarkan hasil yang diperoleh dalam penelitian ini, Algoritma ini bekerja dengan cara mencari k tetangga terdekat dari data yang ingin diprediksi dan menentukan kelas mayoritas berdasarkan tetangga tersebut. K-NN dengan nilai k=11 memberikan akurasi sebesar 96.67% dalam mengklasifikasikan spesies bunga Iris.

Tabel 1 :Akurasi

Algoritma	Akurasi (%)
K-Nearest Neighbors (K-NN)	96

### 3.12 Evaluasi Kinerja Model

Setelah menerapkan algoritma klasifikasi, kinerja model dievaluasi menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Tujuannya adalah memastikan model mampu memberikan prediksi yang akurat dan andal. Hasilnya, k-NN menunjukkan precision dan recall yang tinggi, terutama dalam memprediksi Iris-setosa, yang memang lebih mudah dibedakan berdasarkan fitur-fiturnya.

Tabel 2 : Metriks

Metrik	KNN
Akurasi	96%
Precision	Tinggi
Recall	Tinggi
F1-score	Tinggi

Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma K-Nearest Neighbors (KNN) paling efektif dalam menganalisis dataset Iris. Temuan ini bermanfaat bagi peneliti dan praktisi yang ingin menerapkan machine learning untuk klasifikasi, terutama pada dataset sederhana seperti Iris yang punya potensi luas di bidang botani dan pertanian. Perlu dicatat, yang diuji adalah performa model klasifikasi, bukan datasetnya—dataset Iris hanya digunakan sebagai alat untuk melatih dan menguji berbagai model.

#### IV. KESIMPULAN

Penelitian ini berhasil menggunakan algoritma K-Nearest Neighbors (k-NN) untuk mengklasifikasikan spesies bunga Iris dengan akurasi tinggi, yaitu 96,67%. Dari analisis dan visualisasi data, kami menemukan bahwa panjang dan lebar petal adalah fitur utama yang membedakan Iris Setosa dari Iris Versicolor dan Iris Virginica. Hasil evaluasi menunjukkan k-NN sangat efektif untuk mengenali spesies yang punya perbedaan morfologi yang jelas. Temuan ini penting untuk pengembangan aplikasi machine learning di bidang botani dan bisa menjadi dasar untuk penelitian klasifikasi spesies dengan dataset lain di masa depan.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Ibu Erna Daniati, M.Kom selaku dosen pengampu mata kuliah yang telah memberikan bimbingan, arahan, dan masukan yang sangat berharga selama proses penyusunan dan penyelesaian jurnal ini. Ucapan terima kasih juga disampaikan kepada semua pihak yang telah mendukung, baik secara langsung maupun tidak langsung, dalam pelaksanaan penelitian ini.

#### DAFTAR PUSTAKA

- [1] H. Jurnal, Z. Rahman, Z. Fatah, and D. Prasetyo, "JURNAL ILMIAH MULTIDISIPLIN ILMU KLASIFIKASI SPESIES BUNGA IRIS MENGGUNAKAN ALGORITMA KLASIFIKASI KNN DI RAPIDMINER," pp. 60–66, doi: 10.69714/0syd5n74.
- [2] H. Jurnal, Z. Rahman, Z. Fatah, and D. Prasetyo, "JURNAL ILMIAH MULTIDISIPLIN ILMU KLASIFIKASI SPESIES BUNGA IRIS MENGGUNAKAN ALGORITMA KLASIFIKASI KNN DI RAPIDMINER," pp. 60–66, doi: 10.69714/0syd5n74.
- [3] N. T. Ujianto, Gunawan, H. Fadillah, A. P. Fanti, A. D. Saputra, and I. G. Ramadhan, "Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis," *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. 4, no. 1, pp. 33–43, Feb. 2025, doi: 10.24246/itexplore.v4i1.2025.pp33-43.
- [4] A. Devi Mutiara Khoirun Nisa, N. Lailatus Sa, L. Fitriyani, A. Nuzula Fitranti, and A. Setiawan, "PENERAPAN MACHINE LEARNING DENGAN K-NEAREST NEIGHBORS (KNN) UNTUK IDENTIFIKASI JENIS BUNGA BERDASARKAN KARAKTERISTIK FISIK."
- [5] "10. Penerapan Metode Support Vector Machine Dalam Klasifikasi Bunga Iris."
- [6] A. Syarif and A. Ramadhanu, "PENGUNAAN PRINCIPAL COMPONENT ANALYSIS (PCA) DAN K-NEAREST NEIGHBORS (KNN) DALAM KLASIFIKASI SAYURAN MENTIMUN, PARE, DAN TERONG," 2025. [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [7] E. M. Geldiev, N. V. Nenkov, and M. M. Petrova, "EXERCISE OF MACHINE LEARNING USING SOME PYTHON TOOLS AND TECHNIQUES," *CBU International Conference Proceedings*, vol. 6, pp. 1062–1070, Sep. 2018, doi: 10.12955/cbup.v6.1295.
- [8] J. Teknologi *et al.*, "PENGENALAN WAJAH MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS DAN TEKNIK PENGABUNGAN CITRA," 2024. [Online]. Available: <https://journalpedia.com/1/index.php/jtpi>

- [9] M. Shelin Sahira, F. Dessela Putri, A. Ristyawan, and E. Daniati, "Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 502 Penggunaan Algoritma K-Nearest Neighbor Untuk Klasifikasi Data Diabetes Pada Wanita," Online, 2024.
- [10] E. Priyanto *et al.*, "Implementasi Metode K-Nearest Neighbor Untuk Memprediksi Kondisi Cuaca Penulis Korespondensi," Online, 2024.
- [11] O. D. Kurnia *et al.*, "Analisis Perbandingan Algoritma Naïve Bayes dengan K-Nearest Neighbor (KNN) Pada Dataset Mobile Price Classification Penulis Korespondensi: Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 1174," Online, 2024.
- [12] M. Reza Muzaki, I. Melinda Nurfajriana, F. Putri Anugerah Ilahi, A. Ristyawan, and E. Daniati, "Implementasi Data Mining dengan Algoritma K-Nearest Neighbors untuk Memprediksi Risiko Diabetes Menggunakan Chatbot Telegram," Online, 2024.
- [13] D. Fitriono, S. A. Wardani, M. Nizar, B. Al Varuq, A. Ristyawan, and E. Daniati, "Perbandingan Metode Algoritma Decission Tree dan K-Nearest Neighbors untuk Memprediksi Kualitas Air yang dapat dikonsumsi," Online, 2024.
- [14] M. Shelin Sahira, F. Dessela Putri, A. Ristyawan, and E. Daniati, "Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 502 Penggunaan Algoritma K-Nearest Neighbor Untuk Klasifikasi Data Diabetes Pada Wanita," Online, 2024.
- [15] "10. Penerapan Metode Support Vector Machine Dalam Klasifikasi Bunga Iris."
- [16] M. Reza Muzaki, I. Melinda Nurfajriana, F. Putri Anugerah Ilahi, A. Ristyawan, and E. Daniati, "Implementasi Data Mining dengan Algoritma K-Nearest Neighbors untuk Memprediksi Risiko Diabetes Menggunakan Chatbot Telegram," Online, 2024.