

Prediksi Jenis Bunga Iris dengan Algoritma K-Nearest Neighbor Menggunakan RapidMiner

Zaehol Fatah¹, Ahmad Fauzan²

^{1,2,3} Teknologi Informasi, Universitas Ibrahimy, Situbondo

Email : ¹zaeholfata@gmail.com, ²fauzanprasetyo0321@gmail.com

Abstrak

Pada kesempatan ini peneliti bertujuan untuk mengklasifikasi tumbuhan Iris Species menggunakan metode algoritma K-Nearest Neighbor yang mengimplementasikan dalam RapidMiner Studio 2025.0.1. Yang mana Dataset berisi 150 entri yang mewakili tiga spesies iris (sclerotium, iris beraneka warna, dan eceng gondok), masing-masing berisi empat fitur utama yang digunakan sebagai variabel studi: ukuran Panjang dan lebar dibagian kelopak serta mahkota bunga. Metodologi penelitian ini melibatkan beberapa Langkah yaitu: akuisisi data (Retrieve Iris), penandaan fitur (Set Role), pemisahan data (70% data pelatihan, 30% data uji), pelatihan model menggunakan algoritma K-NN dengan K = 5 dan pembobotan, implementasi model (Apply Model), dan evaluasi kinerja. Hasil pengujian menunjukkan bahwa model K-NN dapat mengklasifikasikan tumbuhan Iridaceae dengan akurasi 97,78%, tingkat presisi dan recall melebihi 93%, dan kesalahan klasifikasi yang sangat rendah. Penelitian ini juga menunjukkan efektifitas tinggi algoritma K-NN dalam pengenalan pola morfologi bunga dan juga menunjukkan bahwa aplikasi RapidMiner dapat dijadikan alat yang efektif dan praktis untuk mengembangkan model klasifikasi berbasis pembelajaran mesin.

Kata Kunci: K-Nearest Neighbor, RapidMiner, Klasifikasi, iris, Pembelajaran mesin.

Abstract

On this occasion, the researcher aims to classify Iris Species plants using the K-Nearest Neighbor algorithm method implemented in RapidMiner Studio 2025.0.1. The dataset contains 150 entries representing three iris species (sclerotium, variegated iris, and water hyacinth), each containing four main features used as study variables: petal length and width, and petal length and width. The methodology of this study involves several steps, namely: data acquisition (Retrieve Iris), feature marking (Set Role), data separation (70% training data, 30% test data), model training using the K-NN algorithm with K = 5 and weighting, model implementation (Apply Model), and performance evaluation. The test results show that the K-NN model can classify Iridaceae plants with an accuracy of 97.78%, precision and recall levels exceeding 93%, and very low classification errors. This study also shows the high effectiveness of the K-NN algorithm in recognizing flower morphological patterns and also shows that the RapidMiner application can be used as an effective and practical tool for developing machine learning-based classification models.

Keywords: K-Nearest Neighbor, RapidMiner, Classification, iris, Machine learning.

PENDAHULUAN

Klasifikasi species merupakan tumbuhan yang berdasarkan ciri morfologi merupakan aplikasi umum pada mesin pembelajaran yang mana untuk mendukung klasifikasi otomatis. Fisher Reference Dataset merupakan salah satu dataset yang

sering sekali digunakan dalam penelitian klasifikasi, yang berisi 150 sampel dan terdiri dari tiga spesies dan empat sifat morfometrik (Ningati & Hanik, 2025).

Dataset ini sudah menyediakan beberapa studi kasus yang sangat baik untuk menguji kemampuan algoritma yang

digunakan pada pembelajaran mesin dalam mengenali pola morfologi yang membedakan kelas spesies (Wulandari, 2024).

Pada kali ini peneliti menggunakan alqoritma K-Nearest Neighbor yang mana sudah sering dipilih untuk tugas klasifikasi dikarenakan mempunyai prinsip sederhana tetapi efektif: Mengklasifikasi data baru sesuai dengan jumlah besar tetangga terdekat dalam ruang fitur, pada metode alqoritma K-NN mempunyai keunggulan tersendiri yaitu tidak memerlukan fase Latihan model yang rumit dan juga mudah untuk diinterpretasikan. Namun, kinerjanya sensitif pada skala sifat, teknik praproses, dan pemilihan parameter k (jumlah tetangga), sehingga masih membutuhkan desain eksperimen yang cermat (Hasan Efendi dkk, 2025).

Lebih lanjut, platform analitik visual seperti RapidMiner telah dijadikan pilihan yang sangat populer untuk penelitian implementasi alqoritma K-NN dikarenakan memberikan akses implementasi alur kerja praproses, pelatihan, dan evaluasi dengan tanpa memerlukan pengembangan perangkat lunak yang ekstensif (Fahmi, 2023).

Para peneliti terdahulu sudah menunjukkan hasil bahwa bervariasi dalam dataset iris tergantung kepada konfigurasi eksperimen. Sementara pada RapidMiner mengimplementasikan alqoritma K-NN dengan akurasi yang sangat tinggi, bahkan beberapa temuan sudah menunjukkan normalisasi dan menemukan nilai k yang optimal diperlukan untuk stabilitas model (Dwiputra et al., 2024; Octavianto & Subtari, 2025).

Dari pandang biologis, fitur morfologi bunga seperti ukuran dan rasio kelopak merupakan factor penting yang nantinya untuk mengetahui perbedaan spesies dalam dataset iris, sehingga pemilihan fitur dapat mempengaruhi keberhasilan klasifikasi (Naufal Hilmi et al., 2024; Ronaldo & Kurnia, 2025). Berdasarkan kreteria ini, peneliti ingin menerapkan alqoritma K-NN pada dataset

iris menggunakan RapidMiner juga menganalisis kinerja model berdasarkan konfigurasi yang dipilih (pemisahan data, nilai k , dan praproses). Secara spesifik, kami (1) mengembangkan diagram alir untuk klasifikasi K-NN di RapidMiner, yang didalamnya termasuk akuisisi data, penetapan peran, partisi data, pelatihan, dan juga evaluasi. (2) mengajukan indicator kinerja utama (akurasi, presisi, recall, dan skor F1) (Beliya et al., 2023; Sari & Sari, 2022). (3) membahas dampak pengaturan parameter pada hasil klasifikasi terkait fitur morfologi iris.

Karya ini berkomitmen dalam menyediakan dokumen praktis eksperimen alqoritma K-NN juga menganalisis kinerja di RapidMiner, dan juga berfungsi sebagai refrensi bagi peneliti pemula yang juga ingin menerapkan metode yang sama pada set data morfologi. Karya ini juga disusun sebagai berikut: setelah pendahuluan ini, pertama-tama kami menjelaskan metodologi penelitian kami, yang didalamnya mengenai sumber data juga alur implementasi di RapidMiner. Kemudian, kami juga menyajikan hasil dan membahas evaluasi model. Pada akhirnya, kami menawarkan kesimpulan juga saran untuk terus dikembangkan lebih lanjut (Ahmad Sabri, 2023).

METODE

Pada studi kali ini menggunakan pendekatan empiris kuantitatif yang mana untuk mengklasifikasi dataset iris menggunakan alqoritma K-NN. Studi ini juga dilakukan menggunakan perangkat lunak RapidMiner Studio versi RapidMiner Studio menyediakan lingkungan alur kerja visual untuk membangun, melatih, dan mengevaluasi pada suatu model pembelajaran mesin tanpa perlu pengkodean manual.

1. Dataset dan Sumber

Dataset yang digunakan dalam studi ini adalah Iris Dataset, yang mencakup 150 spesimen bunga dari tiga spesies: iris keras, iris multiwarna, dan iris biasa. Dataset ini menyediakan empat karakteristik numerik

utama (panjang dan lebar batang, serta panjang dan lebar kelopak) yang digunakan sebagai variabel studi. Dataset ini tersedia untuk umum melalui Repositori Pembelajaran Mesin Universitas California, Irvine, yang memungkinkan para peneliti untuk mengaksesnya tanpa batasan lisensi dan menggunakannya untuk mengembangkan model klasifikasi. Dataset ini dianggap sebagai standar internasional dan sering digunakan sebagai referensi untuk eksperimen pembelajaran mesin karena struktur datanya yang jelas dan seimbang serta pentingnya pengujian algoritma klasifikasi berdasarkan pola bentuk bunga. (Yana & Nafi'iyah, 2021).

2. Alur Studi



Gambar 1 Alur Penelitian Klasifikasi Iris Menggunakan Algoritma K-Nearest Neighbor di RapidMiner

Alur pencarian pada Gambar 1 mengilustrasikan langkah-langkah yang terlibat dalam pemrosesan data RapidMiner. Ini meliputi pengumpulan data (pengumpulan irisan), penetapan peran properti (penetapan peran), segmentasi data menjadi data pelatihan dan uji (segmentasi data), pelatihan model menggunakan algoritma K-NN, penerapan model ke data uji (aplikasi model), dan evaluasi kinerja model menggunakan pengontrol kinerja. Alur ini menjalankan proses klasifikasi secara sistematis dan metodis. Yang mana datasetnya berisi 150 titik data dengan lima atribut (empat atribut prediktif dan satu label kelas) (Salsabila et al., 2021). Selanjutnya yaitu tahap penetapan peran yang dilakukan untuk menentukan atribut target yang akan dijadikan sebagai label pengambilan. Dalam hal ini, atribut tipe ditetapkan sebagai kelas. Selanjutnya yaitu dilakukanlah pemisahan data, Pisahkan

pada datasetnya jadi dua macam: 70% untuk data latih dan 30% untuk data uji. Pada bagian ini secara acak yang gunanya untuk menjaga distribusi data yang seimbang (pengambilan sampel acak) (Farokhah, 2020). Pada tahap selanjutnya adalah peran dari algoritma K-NN dengan parameter $k = 5$ dan metode pembobotan. Algoritma ini akan mengkalkulasikan Euclidean dari data yang uji daan juga data latih dan juga menentukan kelas berdasarkan mayoritas tetangga terdekat. Model alqortma K-NN yang telah dibangun untuk diterapkan menggunakan operator “Terapan Model” yang mana untuk menghasilkan prediksi label dan menampilkan tingkat keyakinan pada setiap kelas yang diprediksi (Neighbors, n.d.). Tahap terakhir yaitu “Performa”, yang mana bertujuan untuk mengevaluasi performa model berdasarkan metrik presisi, keyakinan, dan perolehan Kembali . Hasil percobaan menunjukkan bahwa model K-NN mencapai tingkat akurasi sebesar 97,78% dengan tingkat presisi dan recall rata-rata melebihi 93%.

3. Parameter Pengujian

Parameter utama yang akan kami gunakan terhadap penelitian ini di tunjukkan pada tabel sebagai berikut.

Tabel 1 Parameter Pengujian pada Algoritma K-Nearest Neighbor

No	Parameter	Nilai	Keterangan
1	Jumlah tetangga (K)	5	Tetangga terdekat
2	Pembobotan	Weighted	Memberi bobot pada jarak terdekat
3	Rasio pembagian data	70:30	70% training, 30% testing
4	Distance Measure	Euclidean Distance	Pengukuran jarak antar data
5	Label	Species	Kelas target

Tabel 1 menunjukkan parameter yang digunakan dalam proses pengujian algoritma K-NN (termasuk nilai k , metode pembobotan, skala jarak yang digunakan, dan rasio distribusi data pelatihan dan uji). Parameter-parameter ini dirancang untuk memastikan model mencapai kinerja yang optimal dan konsisten selama pelatihan dan evaluasi.

HASIL DAN PEMBAHASAN

Bagian ini menyajikan temuan penelitian, mulai dari pengumpulan data dan prapemrosesan hingga pengembangan dan evaluasi kinerja model K-NN, berdasarkan metodologi yang dijelaskan di atas.

1. Pengumpulan Data

Kumpulan data yang digunakan terdiri dari 150 sampel yang mewakili tiga spesies iris (Iris setosa, Iris versicolor, dan Iris virginica). Setiap sampel berisi empat atribut numerik: panjang kelopak, lebar kelopak, panjang kelopak, dan lebar kelopak. Data ini menjadi dasar untuk proses klasifikasi menggunakan algoritma K-NN.

2. Prapemrosesan Data

Pada langkah prapemrosesan, atribut "jenis" ditentukan menggunakan operator Set Role. Atribut "jenis" mewakili kategori klasifikasi, dan empat atribut lainnya disebut atribut yang diharapkan. Parameter partisi data kemudian digunakan untuk membagi kumpulan data menjadi dua subset: 70% untuk data latih dan 30% untuk data uji. Partisi acak ini bertujuan untuk mempertahankan distribusi data dan membuat evaluasi model yang mewakili data yang sebelumnya tidak diproses oleh model.

3. Pengembangan Model K-NN

Model klasifikasi ini dikembangkan menggunakan koefisien K-NN $k=5$ dan metode pembobotan. Pada langkah ini, algoritma menghitung jarak Euclidean antara data uji dan data latih untuk menentukan kelas mayoritas tetangga terdekat. Model berhasil mempelajari pola tipikal untuk tiga kelas bunga berdasarkan empat atribut numerik.

4. Pengujian dan Performa Model

Model yang dihasilkan diaplikasikan pada data uji menggunakan koefisien aplikasi model. Hasil prediksi menunjukkan kelas sebenarnya, kelas prediksi, dan tingkat keyakinan model untuk setiap prediksi.

Kelas	Precision (%)	Recall (%)
Iris-setosa	100.00	93.33
Iris-versicolor	93.75	100.00
Iris-virginica	100.00	100.00

Evaluasi dilakukan menggunakan koefisien performa, yang secara otomatis menghitung akurasi, tingkat kecocokan, keamanan, dan matriks konfusi. Hasil pengujian menunjukkan bahwa algoritma K-NN menggunakan $k=5$ dan koefisien pembobotan mencapai akurasi 97,78%, dengan akurasi dan keamanan melebihi 93%. Kesalahan klasifikasi minimal, dengan hanya satu kesalahan klasifikasi pada genus *Ayame*. Hasil ini konsisten dengan penelitian sebelumnya yang menunjukkan efisiensi tinggi algoritma K-NN pada dataset *Ayame* menggunakan RapidMiner dan Python. Oleh karena itu, model ini sangat cocok untuk diaplikasikan pada dataset bentuk kelopak, karena kestabilannya, keakuratannya yang tinggi, dan struktur khasnya yang jelas.

SIMPULAN

Berdasarkan dengan hasil dari penelitian kami bahwa sanya bisa kami simpulkan alqoritma K-NN pada dataset Iris dengan menggunakan perangkat lunak RapidMiner Studio 2025.0.1 menghasilkan bahwa hasil klasifikasi yang sangat baik. Dengan inialisasi parameter $K = 5$, dengan metode pembobotan tertimbang, dan rasio pembagian data untuk pelatihan 70% dan untuk pengujian 30% model mencapai akurasi 97,78% dengan peresisi dan recall tinggi di tiga kelas bunga. Hasil ini menunjukkan betapa efektivitas alqoritma K-NN dalam mengenali pola morfologi klasifikasi bunga Iris dilakukan dengan menggunakan empat variabel numerik utama, yaitu ukuran panjang dan lebar pada sepal serta petal. Model ini juga menunjukkan kinerja yang stabil di seluruh kelas dengan tingkat kesalahan yang sangat minim dan juga hasil ini mendukung penelitian sebelumnya (Rahman dkk, 2024; Efendi dkk, 2025) yang telah menunjukkan akurasi yang sangat tinggi dalam klasifikasi data biologis, terutama pada dataset dengan struktur fitur yang terukur dengan jelas.

Lebih lanjut, pembelajaran ini menunjukkan bahwa RapidMiner memfasilitasi konstruksi model klasifikasi secara visual dan sistematis tanpa perlu pemrograman manual. Hal ini menjadikan aplikasi RapidMiner alternatif yang efektif bagi mahasiswa dan peneliti bagi pemula yang terlibat dalam analisis data yang berbasis mesin. Lebih lanjut lagi, disarankan untuk membandingkan model ini dengan algoritma lain, seperti pohon keputusan, Naïve Bayes, dan Support Vector Machine (SVM), yang memiliki tujuan untuk pemahaman yang lebih mendalam tentang kinerja pada dataset yang sama.

UCAPAN TERIMA KASIH

Para penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada program penelitian Teknologi Informasi di Universitas Ibrahimy, Situbondo, atas dukungan dan fasilitas penelitiannya dan juga saya ingin mengucapkan banyak banyak terima kasih pada dosen pembimbing kami yang selalu mengarahkan kami, kontribusi, dan saran selama kami mengerjakan jurnal ini. Dan kami juga mengucapkan banyak terimakasih terhadap teman-teman yang telah memrikan tenaga dalam eksperimen dan validasi dalam menggunakan perangkat RapidMiner, serta kepada teman-teman saya yang telah rela memberikan dukungannya baik dukungan moral dan teknis bagi keberhasilan meneliti. Saya sebagai penulis sangat menginginkan dari hasil penelitian ini bisa bermanfaat dan dapat menjadi referensi bagi peneliti yang lain yang mengembangkan metode klasifikasi berbasis pembelajaran mesin di bidang ilmu computer dan data.

DAFTAR PUSTAKA

Ahmad Sabri. (2023). Transfer Learning Model Cnn Pralatih Untuk Klasifikasi Bunga Iris Berbasis Citra. *Seminar Nasional Teknologi Informasi Dan Komunikasi STI&K (SeNTIK)*, 7(1), 2581–2327.

- Beliya, S., Syarif, T. R., Afian, R. N., Afrilia, U. L., & Putra, R. B. (2023). Jurnal ilmiah Multidisiplin Ilmu Nusantara Jurnal ilmiah Multidisiplin Ilmu Nusantara. *Jurnal Ilmiah Multidisiplin Ilmu Nusantara*, 1(1), 14–19.
- Dwiputra, M., Wijaya, I., & Dwiyanaputra, R. (2024). Perancangan Mesin Klasifikasi Menggunakan Particle Swarm Optimization Designing A Classification Machine Using Particle Swarm Optimization. *J-COSINE (Journal of Computer Science and Informatics Engineering)*, 8(2), 146–154. <http://jcosine.if.unram.ac.id/>
- Fahmi, M. N. (2023). Implementasi Machine Learning menggunakan Python Library: Scikit-Learn (Supervised dan Unsupervised Learning). *Sains Data Jurnal Studi Matematika Dan Teknologi*, 1(2), 87–96. <https://doi.org/10.52620/sainsdata.v1i2.31>
- Farokhah, L. (2020). Implementasi K-Nearest Neighbor Untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna Rgb Implementation of K-Nearest Neighbor for Flower Classification With Extraction of Rgb Color Features. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 7(6), 1129–1136. <https://doi.org/10.25126/jtiik.202072608>
- Naufal Hilmi, A., Yulia Puspaningrum, E., Endah Wahanani, H., Rungkut Madya No, J., Anyar, G., Gn Anyar, K., & Timur, J. (2024). Implementasi Algoritma K-Nearest Neighbor (KNN) untuk Identifikasi Penyakit pada Tanaman Jeruk Berdasarkan Citra Daun. *Router : Jurnal Teknik Informatika Dan Terapan*, 2(2), 107–117.
- Neighbors, M. A. K. (n.d.). *Analisis Klasifikasi Spesies Bunga Iris. I*, 1798–1804.
- Ningati, R. K., & Hanik, N. R. (2025).

- Identification of Ornamental Flower Plants in The Blooms Garden Bali as a Learning Resource for High School Biology. *Jurnal Biologi Tropis*, 25(1), 402–415.
<https://doi.org/10.29303/jbt.v25i1.8508>
- Octavianto, M. D. A., & Subtari, T. (2025). Analisis Perbandingan Kinerja Algoritma Klasifikasi Data Menggunakan Metode K-NN , Naive Bayes , dan Decision Tree pada Dataset UCI Iris. *MEANS (Media Informasi Analisa Dan Sistem)*, 10(1).
- Ronaldo, R., & Kurnia, Y. (2025). Perbandingan Kinerja Algoritma SVM, Decision Tree, dan, Naive Bayes untuk Klasifikasi dan Pengelompokan Spesies Iris. *POTERS (Proceedings of Technology, Engineering and Computers)*, 1(1), 70–77.
- Salsabila, A., Yunita, R., & Rozikin, C. (2021). Identifikasi Citra Jenis Bunga menggunakan Algoritma KNN dengan Ekstrasi Warna HSV dan Tekstur GLCM. *Technomedia Journal*, 6(1), 124–137.
<https://doi.org/10.33050/tmj.v6i1.1667>
- Sari, W. S., & Sari, C. A. (2022). Klasifikasi Bunga Mawar Menggunakan Knn Dan Ekstraksi Fitur Glcm Dan Hsv. *Skanika*, 5(2), 145–156.
<https://doi.org/10.36080/skanika.v5i2.2951>
- Wulandari, A. (2024). Profil Morfologi Bunga Angiospermae di Lingkungan Fakultas Ilmu Tarbiyah dan Keguruan Universitas Islam Negeri Sunan Kalijaga sebagai Sumber Belajar *Jurnal Exact: Kajian Kemahasiswaan*, 2(1), 88–97. <https://ejournal.uin-suka.ac.id/tarbiyah/exact/article/view/8954>
<https://ejournal.uin-suka.ac.id/tarbiyah/exact/article/download/8954/3608>
- Yana, Y. E., & Nafi'iyah, N. (2021). Klasifikasi Jenis Pisang Berdasarkan Fitur Warna, Tekstur, Bentuk Citra Menggunakan SVM dan KNN. *RESEARCH: Journal of Computer, Information System & Technology Management*, 4(1), 28.
<https://doi.org/10.25273/research.v4i1.6687>