Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 5 No. 2 (2026) pp: 12948-12956
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Klasifikasi Bunga Iris menggunakan Algoritma K-Nearest Neighbor (KNN) Classifier

Enjelina Sukamto[1], Dwi Titi Maesaroh[2], Hari Purwadi[3]
[123]Jurusan Teknologi Informasi, Program Studi Teknik Komputer, Politeknik Negeri Samarinda
[1]enjellina0707@gmail.com, [2]dwititi.maesaroh@polnes.ac.id, [3]hari.purwa06@gmail.com

Abstrak

Identifikasi taksonomi tumbuhan seringkali sulit dilakukan karena adanya unsur subjektivitas dan ketidak efisienan waktu. Pendekatan komputasional yang objektif untuk membedakan spesies berdasarkan ciri morfologis dapat diwujudkan melalui penerapan algoritma pembelajaran mesin, khususnya klasifikasi berbasis contoh. Dengan menggunakan dataset publik Iris yang bersumber dari UCI Machine Learning Repository standar yang terdiri dari 150 sampel, penelitian ini menyelidiki penggunaan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan tiga jenis bunga Iris (Setosa, Versicolor, dan Virginica). Empat parameter dimensi panjang dan lebar sepal serta kelopak menjadi fokus utama pengamatan. Dalam penelitian ini, ditemukan bahwa spesies Setosa memiliki batas spasial yang berbeda secara linier, sedangkan kelas Versicolor dan Virginica memiliki zona tumpang tindih yang signifikan dalam ruang fitur kelopak. Hasil ini diperoleh dengan menggunakan berbagai teknik Analisis Data Eksploratori (EDA), termasuk statistik deskriptif, histogram, dan pemetaan diagram sebar. Untuk pemodelan, data dibagi menjadi 80% data pelatihan dan 20% data uji dengan menggunakan stratifikasi guna menjaga keseimbangan kelas. Parameter lingkungan (neighborhood) ditetapkan sebesar K=5, dan algoritma KNN dievaluasi menggunakan matriks kebingungan serta metrik kinerja klasifikasi. Hasil pengujian menunjukkan bahwa model tersebut memiliki akurasi global sebesar 96,67%. Analisis tersebut menunjukkan bahwa spesies Setosa memiliki nilai presisi dan recall total sebesar 1,00 pada tingkat kelas; namun, terdapat satu kesalahan klasifikasi pada kelas Virginica yang seharusnya diklasifikasikan sebagai Versicolor. Hal ini membuktikan adanya tantangan tumpang tindih fitur alami akibat kemiripan morfologis yang sangat identik antara kedua spesies tersebut.

Kata Kunci: *Machine Learning, Klasifikasi, Bunga Iris, K-Nearest Neighbor, Confusion Matrix, Morfologi Tumbuhan.*

1. Latar Belakang

Berbagai paradigma penelitian dalam bidang sains empiris telah diubah oleh perkembangan kecerdasan buatan. Sistem komputasi otomatis beralih dari metode observasi manual. Machine Learning atau pembelajaran mesin, menjadi alat penting dalam payung kecerdasan buatan untuk analitik. Ini memungkinkan sistem komputer untuk menemukan pola kompleks dalam kumpulan data tanpa memerlukan pemrograman aturan yang ketat. Kemampuan ini sangat penting dalam bidang biologi dan botani. Secara historis, identifikasi spesies tanaman didasarkan pada pengamatan morfologi fisik oleh ahli taksonomi; proses ini rentan terhadap kesalahan manusia dan membutuhkan keahlian khusus yang khusus (Witten dkk., 2021). Oleh karena itu, digitalisasi taksonomi menggunakan algoritma klasifikasi telah menjadi fokus penelitian saat ini. Penelitian ini bertujuan untuk membangun sistem pengenalan taksonomi yang mampu mengklasifikasikan tiga spesies bunga Iris (Iris Setosa, Iris Versicolor, dan Iris Virginica) secara cepat dan presisi menggunakan implementasi algoritma *K-Nearest Neighbor* (KNN).

Dataset bunga Iris, pertama kali diperkenalkan oleh ahli statistik dan biologi R.A. Fisher pada tahun 1930-an, menyimpan pengukuran morfologi dari tiga spesies endemik Iris setosa, Iris versicolor, dan Iris virginica. Dataset ini masih digunakan sebagai tolok ukur (benchmark) dalam pengujian algoritma klasifikasi pola. Empat variabel independen, panjang sepal, lebar sepal, panjang mahkota (petal), dan lebar mahkota, dicatat dalam dataset ini dalam satuan sentimeter (Müller). Meskipun ukurannya relatif kecil (150 sampel), Dataset Iris secara efektif merepresentasikan kompleksitas permasalahan klasifikasi pada data riil. Disparitas tingkat kesulitan pemisahan antarkelas merupakan masalah utama yang tersimpan dalam data ini. Karakteristik ukuran spesies Setosa membuatnya membentuk klaster yang terisolasi secara tegas di dalam ruang multidimensional. Namun, rintangan analitik yang sebenarnya terletak pada pemisahan antara kelas Versicolor dan Virginica. Pada titik ekstrem pertumbuhannya, kedua spesies ini memiliki dimensi fisik yang sangat identik. Akibatnya, terjadi

tumpang tindih (*overlapping*) data yang menuntut algoritma klasifikasi untuk sangat peka dalam menarik batas keputusan (*decision boundary*) (Geron, 2022).

Akibat kesederhanaan logikanya dan ketangguhannya dalam menangani batas kelas non-linear, algoritma klasifikasi K-Nearest Neighbor (KNN) seringkali menjadi opsi rasional untuk mengurai kompleksitas data spasial semacam ini. KNN menggunakan pendekatan lalai atau instance-based learning. Ini membedakannya dari metode parametrik seperti regresi logistik, yang berusaha membuat garis matematis global. Metode ini memetakan seluruh data yang dilatih ke dalam ruang geometri multivariat. KNN menentukan label kelasnya ketika sampel baru, atau data uji, ditambahkan. Ini dilakukan berdasarkan proksimitas jarak terhadap sejumlah **K** titik terdekat (tetangga) di ruang tersebut (Alpaydin, 2020). Meskipun mekanismenya tampaknya sederhana, keberhasilan KNN sangat bergantung pada analisis eksplorasi data yang tepat dan pemilihan parameter hiper (hyperparameter) K yang tepat.

Tujuan penelitian ini adalah untuk mempelajari secara menyeluruh kemampuan algoritma KNN untuk mengklasifikasikan spesies bunga Iris tanpa menggunakan proses standardisasi buatan. Kajian ini tidak hanya bertujuan untuk mencapai angka akurasi yang tinggi, tetapi juga membedah proses klasifikasi itu sendiri. Ini dimulai dengan analisis distribusi statistik, menggambarkan korelasi antar fitur melalui visualisasi grafis berlapis, dan melakukan evaluasi kritis terhadap matriks kebingungan (confusion matrix) pada titik parameter **K**=5. Diharapkan, melalui paparan yang sistematis, makalah ini dapat memberikan justifikasi empiris mengenai batasan dan potensi arsitektur.

2. Metode Penelitian

Kerangka operasional eksperimen ini disusun menggunakan bahasa pemrograman Python, yang dijalankan pada platform analisis data *science*, sepenuhnya memudahkan arsitektur komputasional. Skrip penelitian menggunakan kerangka kerja(*framework*) berbasis pustaka industri yang umum untuk memastikan validitas dan skalabilitas: Numpy dan Pandas mengelola struktur manipulasi matriks, *Matplotlib* dan *Seaborn* memberikan perenderan visualisasi grafik dua dimensi, dan modul *Scikit-Learn*, yang dipercaya, digunakan untuk menginject logika algoritma pembelajaran mesin. Sebagai contoh, rangkaian metode direkayasa ke dalam beberapa sekuens berurutan.

a. Akuisisi dan Pemahaman Himpunan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder publik yang pada dasarnya bersumber dari pangkalan data *UCI Machine Learning Repository*. Dataset Iris yang legendaris ini pertama kali dikumpulkan oleh ahli botani Edgar Anderson dan dipublikasikan secara matematis oleh Ronald Fisher pada tahun 1936. Dataset ini memuat 150 sampel spesimen yang terbagi secara merata ke dalam tiga kelas taksonomi (masing-masing 50 sampel) yaitu Iris setosa, Iris versicolor, dan Iris virginica.

Secara teknis, proses akuisisi data awal dilakukan karena dataset tersebut telah tertanam secara *native* di dalam modul sklearn.datasets. Setelah data dimuat, struktur DataFrame berbasis Pandas digunakan untuk mengubah kolom dan mengekstrak indeks. Untuk variabel independen (fitur morfologi), dataset ini mengintegrasikan struktur matriks berukuran 150 x 4. Sementara itu, variabel target direpresentasikan dalam larik satu dimensi yang mengandung representasi numerik kelas (0 untuk *Setosa*, 1 untuk *Versicolor*, dan 2 untuk *Virginica*). Untuk mempermudah interpretasi visual dan analitik tabular di tahap akhir, pemetaan *string* (*mapping*) digunakan untuk menerjemahkan nilai integer target tersebut kembali ke nama spesies aslinya.

b. Analisis Data Eksploratif (Exploratory Data Analysis)

Bagaimana data tersebar di dalam ruang n-dimensi memengaruhi keakuratan absolut klasifikasi jarak. Analisis Data Eksploratif (EDA) dilakukan dengan sangat hati-hati untuk mengungkap struktur tersembunyi distribusi observasi. Proses ini dirancang untuk lebih dari sekedar pelaporan statistik. Sebaliknya, itu dimaksudkan sebagai fungsi audit untuk menemukan ketimpangan varians yang dapat mengganggu perhitungan geometris. Metode EDA mencakup:

1) **Ekstraksi Statistik Deskriptif:** Dengan menggunakan fungsi agregasi Pandas, nilai ekstrem (minimum dan maksimum), nilai rerata (mean), dan standar deviasi tersebar di setiap dimensi kelopak dan sepal.
2) **Pemetaan Distribusi Univariat (Histogram):** membuat matriks grafik batang untuk melihat apakah bentuk distribusi fitur menunjukkan distribusi normal Gaussian atau mengandung skewness.

DOI: https://doi.org/10.31004/riggs.v5i2.11258
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3) **Pemetaan Distribusi Bivariat (Scatter Plot):** Merancang grafik sebaran dua dimensi khusus untuk persilangan karakteristik kelopak, seperti panjang kelopak pada absis dan lebar kelopak pada ordinat. Untuk mengidentifikasi klaster atau tanda tumpang tindih spasial di antara kelas target, representasi warna digunakan secara eksplisit. Palet warna merah, pink, dan magenta digunakan.

4) **Matriks Relasi Multivariat (Pairplot):** Bentuk probabilitas ditampilkan melalui pengeplotan berlapis menggunakan fungsi estimasi kepadatan kernel (KDE) pada diagonal utama. Untuk seluruh kombinasi dimensi morfologi, matriks korelasi scatter digunakan dengan palet warna "Dark2".

c. Partisi Data dan Stratifikasi

Seluruh populasi sampel dibagi menjadi domain pelatihan (training set) dan domain pengujian (testing set) untuk menghindari bias optimisme dalam evaluasi kinerja. Modul train_test_split dari SciKit-Learn dieksekusi dengan mendefinisikan porsi test_size=0.2. Ini juga mengalokasikan 120 observasi sebagai ruang untuk memilih titik referensi KNN, dan 30 observasi dikarantina sebagai proksi data dunia nyata. Untuk memastikan reproduksibilitas eksperimen, parameter isolasi acak ditetapkan pada random_state=46. Untuk tujuan yang lebih penting, parameter stratify=y diaktifkan untuk memastikan proporsi kelas target pada subset uji identik dengan populasi asli. Ini memungkinkan pembagian absolut 10 Setosa, 10 Versicolor, dan 10 Virginica dalam set uji.

```
# Pembagian Dataset
X = iris.data
y = iris.target
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=46, stratify=y)
```

Gambar 1. Training Data

d. Inisialisasi dan Pelatihan Model KNN

Arsitektur klasifikasi utama penelitian ini dikembangkan menggunakan kelas KNeighborsClassifier yang terintegrasi dalam pustaka komputasi SciKit-Learn. K=5, dengan n neighbor=5, adalah setting definitif untuk parameter resolusi ketetanggan. Pilihan bilangan bulat ganjil tingkat menengah ini adalah strategi optimasi heuristik. Nilai 5 diberikan karena secara efektif menghilangkan risiko kebuntuan konsensus suara (*tie-breaking*) saat algoritma harus mengambil keputusan di batas dua kelas. Nilai juga cukup untuk meredam gangguan derau (*noise*) lokal tanpa berisiko menelan batas keputusan global (*underfitting*).

Pengaturan default model adalah metrik Minkowski dengan parameter derajat P=2 , yang secara matematis sama dengan fungsi jarak Euclidean untuk menghitung kedekatan spasial antar vektor. Fungsi komputasi fit digunakan untuk melakukan proses penyesuaian model. Dalam fase ini, seluruh vektor data latihan diproyeksikan secara langsung ke dalam memori internal algoritma tanpa melalui fase penyeragaman skala seperti Normalisasi Skor atau Skala *Min-Max*. pendekatan empiris yang disengaja digunakan untuk membuat keputusan metodologis tentang menginjeksikan fitur mentah (*raw features*) ini. Tujuannya adalah untuk melakukan uji dan evaluasi ketahanan absolut algoritma klasifikasi berbasis instans terhadap variasi magnitudo dimensi alami yang menggambarkan karakteristik spesimen biologis dalam kehidupan nyata.

e. Pengukuran Performa dan Metrik Evaluasi

Prediksi dibuat berdasarkan 30 vektor uji. Fungsi klasifikasi-laporan digunakan untuk mensimulasikan perbedaan antara nilai prediksi dan label faktual (asli). Fungsi ini juga memberikan ringkasan rinci tentang metrik presisi (*precision*), sensitivitas (*recall*), akurasi komprehensif (*accuracy*), dan komposit harmoni (*f1-score*). Untuk menunjukkan lokasi tepat matriks letak *False Positives* dan *False Negatives*, *Confusion Matrix* dibuat dan divisualisasikan dalam bentuk peta panas (*heatmap*) anotasi bergradasi hijau sebagai instrumen validasi akhir. Performa komputasional algoritma dijabarkan melalui analisis persamaan presisi , *recall*, hingga diperoleh *F1-Score*, yang diformulasikan sebagai berikut:

1) Presisi (Precision): Mengkuantifikasi dari seluruh prediksi positif yang dijatuhkan model, seberapa besar rasio akurasinya yang secara faktual memang benar positif.

$$Precision = \frac{TP}{TP+FP}$$

2) Sensitivitas (Recall): Mengukur ketajaman algoritma dalam mencari dan mengidentifikasi keseluruhan objek positif sesungguhnya dalam sebuah kelas tanpa ada yang lolos menjadi *false negative*.

$$Recall = \frac{TP}{TP+FN}$$

DOI: https://doi.org/10.31004/riggs.v5i2.11258
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3) Akurasi Global (Accuracy): Menghitung proporsi global dari semua identifikasi vektor observasi yang secara empiris terbukti akurat.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Untuk melengkapi kuantifikasi saklar tersebut, tabel *Confunsion Matrix* divisualisasikan dalam perenderan gradasi peta panas (*heatmap*). Tabulasi visual matriks kontigensi silang ini bertindak sebagai mikroskop analitik yang akan memperlihatkan di celah kelas transisi mana algoritma KNN tersandung dan memicu positif palsu maupun negatif palsu.

3. Hasil dan Diskusi

Hasil pengujian eksperimental arsitektur *machine learning* dibahas secara menyeluruh di bab ini. Rangkaian komputasi menghasilkan sekumpulan keluaran skalar metrik dan visualisasi grafis yang memberikan gambaran luas tentang perilaku data biologis dalam ruang fitur. Hasil penelitian dianalisis secara sistematis dan kronologis. Ini dimulai dengan tinjauan statistik deskriptif dan Analisis Data Eksploratif (EDA) untuk memetakan distribusi sebaran data dan menemukan area tumpang tindih (*overlapping*) antar kelas taksonomi. Selanjutnya, diskusi berfokus pada evaluasi performa prediktif algoritma *K-Nearest Neighbor* (KNN). Tidak hanya persentase akurasi global yang digunakan untuk mengukur kinerja klasifikator akhir ini, tetapi juga metrik evaluasi mendalam seperti presisi, sensitivitas (*recall*), rata-rata harmonik (F1-Score), dan validasi spasial melalui matriks kebingungan (*confusion matrix*).

a. Dekonstruksi Karakteristik Statistik dan Distribusi Univariat

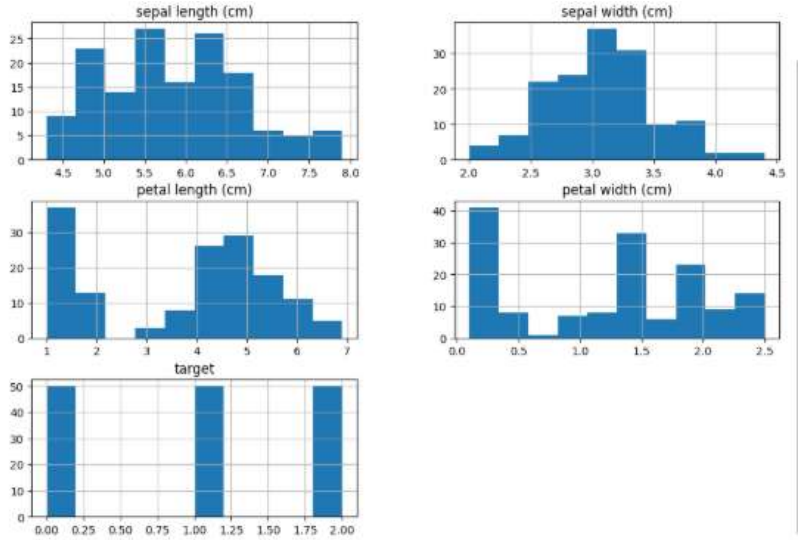
Nilai deskriptif pusat diperiksa sebagai langkah pertama untuk memahami keterbatasan klasifikasi. Parameter morfologi bunga Iris tidak sangat berbeda, menurut tabel agregasi awal. Sepal terpanjang berukuran 7.9 cm dan lebar petal terkecil 0.1 cm. Dispersi nilai menunjukkan perbedaan.

Tabel 1. Ekstraksi metrik statistik deskriptif pada empat fitur dimensi morfologi dataset bunga Iris

Metrik Statistik	Sepal length (cm)	Sepal Width (cm)	Petal length (cm)	Petal Width (cm)
Count	150.00	150.00	150.00	150.00
Mean	5.84	3.05	3.76	1.20
Std dev	0.83	0.43	1.76	0.76
Minimum	4.30	2.00	1.00	0.10
25% (Q1)	5.10	2.80	1.60	0.30
50% (Median)	5.80	3.00	4.35	1.30
75% (Q3)	6.40	3.30	5.10	1.80
Maximum	7.90	4.40	6.90	2.50

Sinyal tersebut diperkuat dengan representasi visual histogram univariat, seperti yang ditunjukkan pada gambar 2. Menurut pengukuran yang terkait dengan arsitektur sepal (sepal length dan sepal width) kurva asimtotik tampak seperti lonceng distribusi normal murni. Di sini, kepadatan probabilitas antar kelas saling tumpang tindih terkonsentrasi pada nilai mean populasi. Struktur distribusi komunal seperti ini menunjukkan bahwa jika dimensi dimensi sepal digunakan sebagai satu-satunya jangkar pembeda identitas spesies, kemampuan diskriminasi deterministiknya sangat lemah.

DOI: https://doi.org/10.31004/riggs.v5i2.11258
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

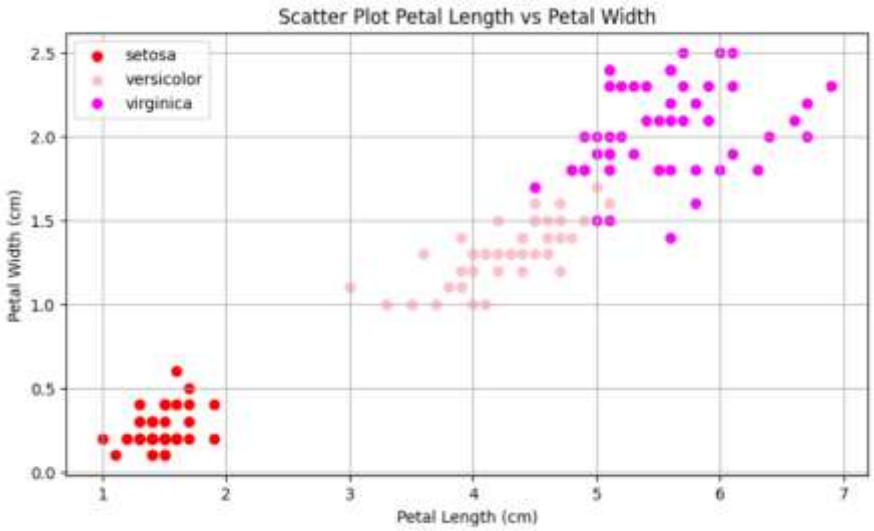


Gambar 2. Visualisasi histogram yang menunjukkan distribusi kepadatan univariat pada masing-masing fitur morfologi bunga Iris.

Sebaliknya, fenomena biomodal dengan dua puncak probabilitas ditunjukkan oleh analisis histogram pada dimensi petal. Dalam panjang petal antara 2.0 dan 3.0 cm dan lebar petal antara 0.6 dan 0.8 cm, terlihat jeda kosong (*gap*). Pemisahan rongga ini menunjukkan anomali morfologi natural di mana satu subpopulasi populasi target memiliki konstruksi genetik yang sangat seragam dan ukurannya lebih kecil daripada populasi target secara keseluruhan.

b. Analisis Ruang Spasial dan Tumpang Tindih Multikelas

Peninjauan Scatter Plot khusus untuk dua fitur petal memvalidasi dan memperkuat penjelasan statistik di atas. Kode warna yang menunjukkan titik sebaran (merah untuk spesies target 0 (Setosa), merah muda untuk spesies 1 (Versicolor), dan magenta pekat untuk spesies 2 (Virginica) menunjukkan batasan penting dalam himpunan data ini (gambar 3).



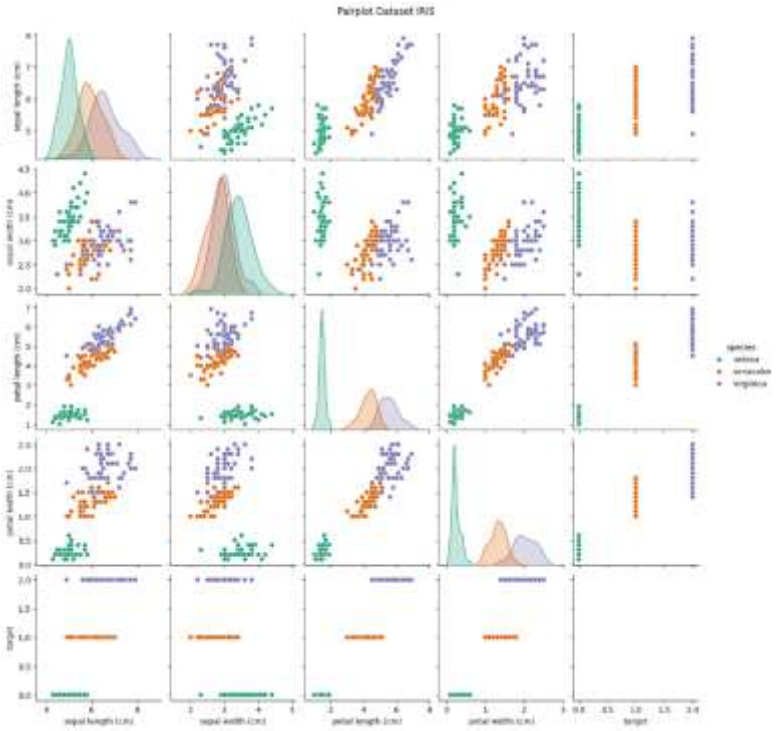
Gambar 3. Pemetaan scatter plot persebaran dimensi ruang fitur antara petal length dan petal width.

Spesimen Setosa (merah) ditemukan di kuadran pojok kiri bawah, membentuk klaster agregat yang sangat terisolasi dan padat. Karena ada margin kekosongan absolut di antara titik terluar Setosa dan perbatasan spesies terdekatnya, fungsi pembelajaran mesin tidak dapat melakukan misklasifikasi di wilayah tersebut.

Tantangan komputasi sesungguhnya muncul pada interaksi antara kelas *Versicolor* (merah muda) dan *Virginica* (magenta). *Versicolor* mendominasi spektrum ukuran menengah, membentang di sepanjang sumbu

DOI: https://doi.org/10.31004/riggs.v5i2.11258
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

kelopak hingga batas atasnya beririsan langsung dengan teritori *Virginica* yang berukuran paling besar. Keduanya berbagi ruang metrik yang sama dan saling tumpang tindih pada rentang panjang petal 4,5 cm hingga 5,5 cm. Area ambiguitas spasial ini sejalan dengan karakteristik dataset biologis yang dikemukakan oleh Efendi et al. (2024), di mana eksplorasi visual menjadi krusial untuk memetakan beban komputasi sebelum model dilatih. Grafik matriks *Pairplot* (Gambar 4) secara holistik mengonfirmasi bahwa irisan serupa juga terjadi secara masif pada dimensi sepal, sehingga melegitimasi pemilihan dimensi petal sebagai prediktor utama.



Gambar 4. Visualisasi data

c. Evaluasi Kinerja Algoritma K-Nearest Neighbor (K=5)

Uji konfrontasi model terhadap 30 data observasi eksternal (menggunakan distribusi acak terstratifikasi) menghasilkan performa prediktif yang sangat impresif. Berdasarkan prinsip mayoritas suara (*plurality voting*) dari lima entitas terdekat (K=5), klasifikasi berbasis jarak ini sukses mencatatkan skor akurasi global sebesar 96,67% (Tabel 2).

Tabel 2. Laporan Klasifikasi dan metrik performa algoritma KNN pada parameter K=5.

Metrik/ Kelas	Presisi (*Presicion*)	Sensitivitas (*Recall*)	*F1-score*	Jumlah Data (*Support*)
Iris Setosa	1.00	1.00	1.00	10
Iris Versicolor	0.91	1.00	0.95	10
Iris Virginica	1.00	0.90	0.95	10
Akurasi Global			0.97 (96,67%)	30
Rata-rata Makro	0.97	0.97	0.97	30
Rata-rata Tertimbang	0.97	0.97	0.97	30

Pencapaian ini mengindikasikan bahwa model mampu meregenerasi aturan klasifikasi dengan margin kesalahan kumulatif kurang dari 4% (Witten dkk., 2021). Temuan ini memberikan kontribusi kebaruan yang signifikan apabila disandingkan dengan penelitian Efendi dkk. (2024). Meskipun penelitian terdahulu tersebut mampu mencapai akurasi identik (96,67%), mereka membutuhkan cakupan parameter ketetanggaan yang jauh lebih luas (K=11). Keberhasilan arsitektur model dalam penelitian ini mencapai stabilitas prediksi yang ekuivalen pada K=5 serta tanpa melalui fase penyeragaman skala standar (*standardization*) membuktikan bahwa

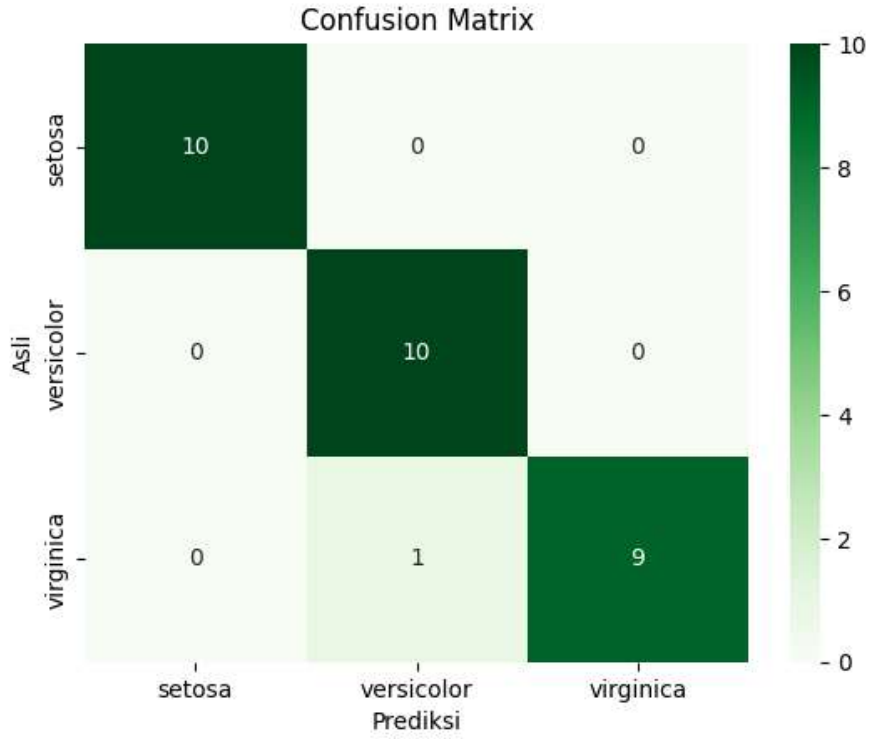
DOI: https://doi.org/10.31004/riggs.v5i2.11258
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

algoritma KNN memiliki efisiensi komputasi yang tinggi dan ketahanan absolut terhadap variasi data mentah asalkan pra-pemrosesan partisinya dilakukan secara tepat.

Laporan klasifikasi (Tabel 2) menunjukkan kinerja asimetris yang merefleksikan realitas morfologis pada pengamatan visual. model mendeteksi kelas Setosa dengan kepastian matematis mutlak tanpa *false positive* maupun *false negative*, menghasilkan skor presisi, *recall*, dan *F1-Score* sempurna di angka 1.00. namun, pengujian di area tumpang tindih memicu dinamika algoritmik. Kategori Versicolor merekam *recall* sempurna (1.00) namun kehilangan sebagian kredibilitas presisi angka 0.91. Sebaliknya, *Virginica* memvalidasi presisi 1,00 absolut namun mencatatkan penurunan sensitivitas menjadi 0,90. Polarisasi metrik di ambang zona persilangan ini merupakan manifestasi logis dari kompetisi sentroid antarkelas di ruang fitur.

d. Analisis Kritis Matriks Kebingungan (Confusion Matrix)

Penyebab fundamental dari fluktuasi skor parsial diatas dijelaskan secara rinci di dalam tabulasi matriks kontingensi (*Confusion Matrix*) bergaya anotasi numerik hijau (Gambar 5).



Gambar 5. Gamheatmap dan confusion matrix hasil prediksi algoritma KNN pada 30 data uji.

Peta panas (heatmap) silang tersebut merekam presisi kategorikal yang memverifikasi diagnosis kerentanan model:

1. Sebanyak 10 data proksi yang memiliki determinasi asli Setosa ditemukan sebagai Setosa di ruang prediksi (TN positif murni).
2. Blok populasi yang awalnya diidentifikasi sebagai Versicolor sukses dibedakan dan diprediksi dengan benar, tanpa terpengaruh oleh prediksi teritori lain.
3. Namun, algoritma mendeteksi salah satu sampel dari sepuluh sampel uji taksonomi Virginica sebagai kelas anomali dengan satu nilai absolut, dan keliru melabelinya menjadi kategori prediksi Versicolor. Di antara sampel yang tersisa, sembilan sampel ditarik dengan benar pada kelas awal.

Sebuah spesimen Virginica dengan label Versicolor adalah salah satu misklasifikasi (classification error) ini. Ini bukan kesalahan matematis dalam fungsi jarak perhitungan, tetapi tanda struktural dari batas biologis. Mengacu pada literatur *machine learning* (Alpaydin, 2020), instans keliru ini kemungkinan merepresentasikan varian *Virginica* kerdil yang memposisikan dirinya sebagai pencilan spasial (*spatial outlier*). Ketika kalkulasi ketetanggaan Euclidean diukur pada parameter K=5 di zona irisan tersebut, vektor mayoritas murni ditempati oleh kerumunan titik *Versicolor*. Sebagai algoritma tipe *lazy-learning*, KNN secara taat tunduk pada aturan

DOI: https://doi.org/10.31004/riggs.v5i2.11258
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

proksimitas spasial dan mengabaikan ketidakpastian alamiah tersebut. Hal ini menyimpulkan bahwa meskipun memiliki efisiensi dan akurasi yang luar biasa tinggi (96,67%), metode algoritmik geometri murni tetap memiliki titik buta (*blind spot*) semantik dalam membedah anomali genetik tanpa adanya tambahan fitur sekunder pendukung di luar dimensi morfologis dasar.

4. Kesimpulan

Studi analitik eksperimental yang dilakukan pada dataset taksonomi bunga Iris membuktikan bahwa klasifikator komputasional *K-Nearest Neighbor* yang diinisialisasi dengan konfigurasi parameter proksimitas spasial K=5 tanpa melalui penyeragaman skala baku menunjukkan tingkat keandalan prediktif yang sangat tinggi. Berdasarkan peninjauan dekonstruktif mendalam (EDA), divergensi dimensi petal berfungsi sebagai katalis pemisah utama yang mendorong algoritma untuk mengisolasi subkelas *Setosa* dengan kepastian absolut (100%). Model menunjukkan kohesi komputasi yang efisien di perbatasan kelas yang terkompresi secara geometris (zona tumpang tindih antara *Versicolor* dan *Virginica*), dengan mencatatkan tingkat akurasi global akumulatif sebesar 96,67%. Deviasi minor yang terisolasi di dalam matriks kontingensi di mana satu spesimen *Virginica* terklasifikasi ke dalam himpunan *Versicolor* merupakan manifestasi dari ambiguitas biologis alami, bukan kegagalan matematis. Keberhasilan model mencapai akurasi tinggi menggunakan injeksi data mentah (*raw data*) mengimplikasikan efisiensi komputasi yang besar, sehingga algoritma ini sangat ideal untuk diimplementasikan pada perangkat bersumber daya rendah sebagai sistem identifikasi botani awal. Meskipun arsitektur model memberikan hasil yang optimal, penelitian ini memiliki keterbatasan fundamental terkait ketergantungan absolut algoritma terhadap jarak geometris dasar. Saat dihadapkan pada pencilan biologis yang mengalami mutasi ukuran, fungsi proksimitas murni rentan memicu misklasifikasi. Selain itu, pengujian masih terbatas pada 150 sampel data statis dengan tingkat keseimbangan kelas yang terdistribusi sempurna. Oleh karena itu, pengembangan penelitian selanjutnya sangat disarankan untuk tidak hanya mengandalkan dimensi fisik, tetapi mengintegrasikan pengayaan fitur tambahan (seperti tekstur daun atau spektroskopi warna). Implementasi fungsi pembobotan jarak adaptif pada algoritma KNN juga direkomendasikan agar spesimen di area batas tidak diklasifikasikan secara buta. Terakhir, model perlu diuji ketahanannya pada himpunan data taksonomi biologis berskala besar yang memiliki distribusi kelas yang tidak seimbang (*imbalanced dataset*).

Referensi

1. Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press. https://mitpress.mit.edu/9780262043793/
2. Arwansyah, A., Susanto, C., & Nurdiansah, N. (2025). Model klasifikasi machine learning berbasis multiple measurement distance. *CSRID (Computer Science Research and Its Development Journal), 17*(2), 256–269. https://doi.org/10.22303/csrid-.17.2.2025.256-269
3. Bahzad, T. C., Abdulazeez, A. M., Zeebaree, D. Q., & Zebari, D. A. (2021). Machine learning classifiers based classification for IRIS recognition. *Qubahan Academic Journal, 1(2), 106–118*. https://doi.org/10.48161/qaj.v1n2a48
4. Efendi, M. H., Pratama, W. S., & Daniati, E. (2024). Analisis Klasifikasi Spesies Bunga Iris Menggunakan Algoritma K-Nearest Neighbors. *INOTEK (Seminar Nasional Inovasi Teknologi)*, 9, 1798-1804.
5. Farahdinna, F., & Shofy, M. N. (2025). Implementasi K-Nearest Neighbor untuk klasifikasi bunga iris. *JATI (Jurnal Mahasiswa Teknik Informatika), 9*(2), 3510-3514. https://doi.org/10.36040/jati.v9i2.13510
6. Farokhah, L. (2020). Implementasi K-Nearest Neighbor untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 7(6), 1129-1136. https://doi.org/10.25126/jtiik.2020722608
7. Geron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media. https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/
8. Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7
9. Hayati, N. (2023). Klasifikasi jenis bunga mawar menggunakan algoritma K-Nearest Neighbour. *Jurnal Informatika Dan Riset, 1*(1), 31-37. https://doi.org/10.36308/iris.v1i1.474
10. KLASIFIKASI SPESIES BUNGA IRIS MENGGUNAKAN ALGORITMA KLASIFIKASI KNN DI RAPIDMINER (Zainur Rahman, Zaehol Fatah, & Jarot Dwi Prasetyo, Trans.). (2024). *Jurnal Ilmiah Multidisiplin Ilmu, 1*(6), 60-66. https://doi.org/10.69714/0syd5n74
11. Kuhn, M., & Johnson, K. (2021). *Applied predictive modeling in biological informatics*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4614-6849-3
12. Müller, A. C., & Guido, S. (2020). *Introduction to machine learning with Python: A guide for data scientists* (2nd ed.). O'Reilly Media. https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/
13. 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), 35–40. https://ejournal.ust.ac.id/index.php/Jurnal_Means/article/view/4982
14. Penerapan Fungsi Exponential Pada Pembobotan Fungsi Jarak Euclidean Algoritma K-Nearest Neighbor. (2022). *Generation Journal, 6*(2), 98-105. https://doi.org/10.29407/gj.v6i2.18070
15. Rahman, B., Fauzi, F., & Amri, S. (2023). Perbandingan hasil klasifikasi data iris menggunakan algoritma K-Nearest Neighbor dan Random Forest. *Journal Of Data Insights, 1*(1), 19–26. https://doi.org/10.26714/jodi.v1i1.135
16. Rahman, Z., Fatah, Z., & Prasetyo, J. D. (2024). Klasifikasi spesies bunga iris menggunakan algoritma klasifikasi KNN di RapidMiner. *Jurnal Ilmiah Multidisiplin Ilmu, 1*(6), 60-66. https://doi.org/10.69714/0syd5n74
17. Taha Chicho, B., Mohsin Abdulazeez, A., Qader Zeebaree, D., & Assad Zebari, D. (2021). Machine Learning Classifiers Based Classification For IRIS Recognition. *Qubahan Academic Journal, 1*(2), 106–118. https://doi.org/10.48161/qaj.v1n2a48

DOI: https://doi.org/10.31004/riggs.v5i2.11258
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

18. Thirunavukkarasu, K., Singh, A. S., Rai, P., & Gupta, S. (2018). Classification of IRIS dataset using classification based KNN algorithm in supervised learning. *2018 International Conference on Computing, Communication and Automation (ICCCA)*. https://doi.org/10.1109/CCAA.2018.8777643
19. Wibowo, A. P., Widiyono, Saifudin, A., Darmawan, S. A., & Budihartono, E. (2022). Naïve Bayes, Neural Network dan K-Nearest Neighbor untuk Klasifikasi Topik Tugas Akhir. *Smart Comp*, 11(4), 661–667. https://doi.org/10.30591/smartcomp.v11i4.4251
20. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2021). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann. https://doi.org/10.1016/C2009-0-19715-5

DOI: https://doi.org/10.31004/riggs.v5i2.11258
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)