

IMPLEMENTASI K-NEAREST NEIGHBOR UNTUK KLASIFIKASI BUNGA IRIS

Frenda Farahdinna, Mochamad Naufal Shofy

Informatika, Universitas Nasional

Jl. Sawo Manila, Pejaten Ps. Minggu Jakarta, Indonesia

frenda.farahdinna@civitas.unas.ac.id

ABSTRAK

Klasifikasi bunga menjadi tantangan dalam bidang botani karena variasi warna dan kemiripan dengan latar belakang, yang dapat menyebabkan kesalahan identifikasi. Kesulitan ini menghambat proses pengelompokan spesies secara akurat, terutama dalam lingkungan dengan karakteristik visual kompleks. Penelitian ini bertujuan untuk mengklasifikasikan tiga spesies bunga Iris menggunakan algoritma *K-Nearest Neighbor* (KNN) berdasarkan fitur morfologi utama, yaitu panjang dan lebar sepal serta petal. Metode yang digunakan mengikuti kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM), mencakup tahapan pemahaman bisnis, pengumpulan dan persiapan data, pemodelan, evaluasi, serta implementasi. Dataset yang digunakan berasal dari kumpulan data bunga Iris yang telah distandarisasi, dan proses klasifikasi dilakukan dengan membandingkan berbagai nilai K untuk menentukan parameter optimal yang meningkatkan akurasi model. Hasil penelitian menunjukkan bahwa algoritma KNN mampu mencapai akurasi hingga 96%, dengan performa terbaik dalam mengidentifikasi spesies Iris setosa. Evaluasi menggunakan *confusion matrix* menunjukkan kesalahan klasifikasi terjadi antara spesies Iris versicolor dan Iris virginica, yang memiliki kemiripan morfologi. Normalisasi data berperan penting dalam meningkatkan stabilitas model dengan mencegah *overfitting* dan *underfitting*. Dengan demikian, algoritma KNN terbukti sebagai metode klasifikasi yang efektif dalam mengelompokkan spesies bunga berdasarkan pola numerik fitur morfologinya.

Kata kunci : Klasifikasi, K-NN, Bunga Iris

1. PENDAHULUAN

Klasifikasi merupakan teknik yang digunakan untuk mengelompokkan objek ke dalam kategori tertentu berdasarkan karakteristik yang dimilikinya. Dalam dunia botani, klasifikasi bunga menjadi tantangan tersendiri karena variasi warna serta kemiripan dengan latar belakang yang dapat menyebabkan kesalahan identifikasi. Angiospermae atau magnoliophyta, sebagai kelompok tanaman berbunga, memiliki struktur reproduksi utama berupa bunga yang berperan dalam proses pembuahan dan penyerbukan. Meskipun berbagai metode telah digunakan dalam klasifikasi bunga, permasalahan utama yang muncul adalah kesulitan dalam membedakan spesies tertentu akibat karakteristik visual yang serupa, terutama dalam lingkungan dengan latar belakang yang kompleks. Selain itu, metode konvensional sering kali membutuhkan keahlian khusus dan proses yang memakan waktu.

Untuk mengatasi permasalahan tersebut, penerapan teknologi *machine learning* menjadi solusi yang efisien dan akurat. Salah satu algoritma yang sering digunakan dalam klasifikasi adalah *K-Nearest Neighbor* (KNN), yang bekerja dengan mencari tetangga terdekat berdasarkan ruang fitur dan menentukan kelas berdasarkan mayoritas tetangga tersebut. Penelitian ini bertujuan untuk mengimplementasikan algoritma KNN dalam klasifikasi tiga spesies bunga Iris guna mengevaluasi efektivitasnya berdasarkan fitur morfologi utama, seperti panjang dan lebar sepal serta petal. Rencana penyelesaian masalah dilakukan dengan pendekatan *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang mencakup pemahaman bisnis

(*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), serta implementasi model (*deployment*) guna mencapai akurasi optimal.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Menurut penelitian yang dilakukan oleh Farokhah, metode klasifikasi K-Nearest Neighbor (KNN) menunjukkan akurasi yang tinggi dalam pengenalan bunga dengan ekstraksi fitur warna RGB, mencapai 95% ketika dibandingkan dengan metode ekstraksi lainnya. Penelitian ini juga mengungkapkan bahwa kombinasi ekstraksi tekstur dan color moment dapat meningkatkan akurasi klasifikasi. Hasil ini menegaskan efektivitas KNN dalam klasifikasi objek berbasis fitur warna dan tekstur [1].

Lebih lanjut, penelitian yang dilakukan oleh Sutri dan Ramadhanu menyoroti bahwa penerapan sistem hybrid PCA-KNN terbukti efektif dalam klasifikasi buah kelapa muda, kelapa tua, dan buah naga. Penelitian ini menggunakan metode Principal Component Analysis (PCA) untuk reduksi dimensi data citra, diikuti dengan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi. Hasil penelitian menunjukkan bahwa kombinasi PCA dan KNN mencapai akurasi tinggi, dengan nilai akurasi rata-rata mencapai 96%, sehingga membuktikan efektivitasnya dalam sistem seleksi buah otomatis [2].

2.2. Data Mining

Data mining merupakan serangkaian prosedur yang dikenal sebagai data mining digunakan untuk

mengeksktraksi pola dan informasi dari suatu basis data untuk mengekstraksi nilai tambahan dari informasi yang sebelumnya tidak diketahui secara manual. Metode ini juga bertujuan untuk mengidentifikasi hubungan antar data serta mengelompokkan data ke dalam satu atau beberapa kategori, di mana objek-objek dalam suatu kelompok memiliki kesamaan yang tinggi di antara mereka [3].

2.3. CRISP-DM

CRISP-DM atau *Cross-Industry Standard Process for Data Mining*, terbagi menjadi enam tahap. CRISP-DM seringkali digunakan sebagai solusi dari akar masalah bisnis [4].

2.4. Machine Learning

Machine learning dapat dikategorikan ke dalam tiga jenis, yaitu *Supervised Learning*, *Unsupervised Learning*, dan *Reinforcement Learning*. Pada jenis *Supervised Learning*, sekumpulan data diberikan untuk mengklasifikasikan kategori yang sebelumnya tidak diketahui. Sebaliknya, dalam pembelajaran tanpa pengawasan, data tidak perlu diberi label, dan prosesnya tidak mencari contoh dari kategori yang sudah ditentukan. Sementara itu, pembelajaran penguatan berada di antara kedua metode tersebut, di mana konsep utamanya adalah mencapai tujuan tanpa instruksi eksplisit dari komputer mengenai pencapaian tersebut [5].

2.5. K-Nearest Neighbor

Algoritma yang tergolong dalam metode *K-Nearest Neighbor* (K-NN) merupakan algoritma terbimbing yang bertujuan untuk mengklasifikasikan objek baru dengan mempertimbangkan atribut dan sampel pelatihan yang ada. Hasil klasifikasi data uji baru ditentukan atas dasar kategori mayoritas dari K tetangga terdekat. Pengklasifikasi dalam metode ini berfungsi dengan menggunakan memori tanpa menggunakan model khusus. Ketika diberikan sebuah titik uji, algoritma ini akan mencari jumlah K terdekat dari objek dengan titik tersebut. Proses klasifikasi dilakukan dengan memilih kategori yang paling banyak muncul di antara K objek terdekat. Algoritma K-NN mengandalkan konsep klasifikasi berdasarkan kedekatan, di mana tetangga-tetangga dekat maupun jauh dikategorikan berdasarkan jarak yang dihitung menggunakan formula Euclidean [6].

3. METODE PENELITIAN

3.1. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah algoritma pembelajaran mesin berbasis instance yang digunakan untuk klasifikasi dan regresi dengan cara membandingkan data baru dengan sejumlah K tetangga terdekat dalam ruang fitur [7].

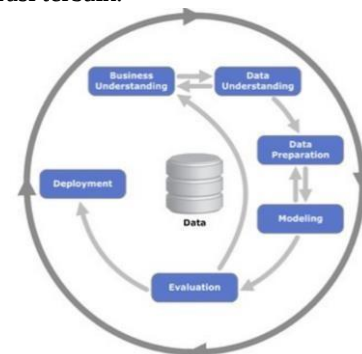
Algoritma ini bekerja dengan menghitung jarak antara sampel data baru dan seluruh data dalam dataset menggunakan metrik seperti Euclidean Distance, lalu menentukan kelas berdasarkan mayoritas kelas dari

tetangga terdekat tersebut. Dalam penelitian yang dijelaskan dalam dokumen, KNN diterapkan untuk klasifikasi tiga spesies bunga Iris (Setosa, Versicolor, dan Virginica) berdasarkan empat fitur morfologi, yaitu panjang dan lebar sepal serta petal. Proses klasifikasi dilakukan dengan membagi dataset menjadi data pelatihan dan data uji, kemudian menormalisasi data untuk mengurangi perbedaan skala fitur, serta memilih nilai K yang optimal guna meningkatkan akurasi model.

3.2. CRISP-DM

Metodologi penelitian yang digunakan dalam penelitian ini adalah penelitian deskriptif dengan pendekatan kuantitatif dan membuat model klasifikasi bunga iris dengan metode K-Nearest Neighbor. Teknik perancangan model mengadopsi metode *Cross Industry Standart Process for Data Mining* (CRISP-DM) mencakup *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation* dan *Deployment* [8].

Hasil penelitian ini diharapkan mendapatkan nilai akurasi terbaik.



Gambar 1. Metode CRISP-DM

Gambar 1 menunjukkan model pengembangan sistem CRISP-DM, yang terdiri dari enam fase sebagai berikut:

- Business Understanding*, fase ini bertujuan untuk memahami kebutuhan, dan membuat perencanaan strategi guna mencapai tujuan.
- Data Understanding*, fase ini dimulai dengan pengumpulan data disertai analisis informasi mendalam. Hal ini termasuk mengidentifikasi masalah kualitas data dan menemukan elemen penting yang bisa digunakan untuk membangun hipotesis mengenai informasi tersembunyi.
- Data Preparation*, fase yang mencakup rangkaian proses untuk menghasilkan dataset akhir, bisa dilakukan secara berulang, mencakup pemilihan tabel, pembersihan data, transformasi data, dan lainnya.
- Modeling*, fase membangun dan melatih model data dengan berbagai teknik. Fase ini memungkinkan proses yang sedang dilakukan kembali pada proses awal.
- Evaluation*, di tahap ini, model yang telah dikembangkan diharapkan memiliki kualitas yang

memadai dari sudut pandang analisis data. Sebelum model digunakan, evaluasi ini berfungsi untuk menentukan apakah model tersebut dapat memenuhi tujuan yang telah ditetapkan dalam Pemahaman Bisnis.

- f. *Deployment*, fase terakhir implementasi hasil model yang bisa diakses untuk digunakan sebagai laporan sederhana dalam pengambilan keputusan.

4. HASIL DAN PEMBAHASAN

4.1. Business Understanding

Klasifikasi bunga atau spesies tanaman lain secara manual sering kali membutuhkan waktu, keahlian, dan data visual yang akurat. Dalam konteks ini, teknologi pembelajaran mesin, khususnya KNN, dapat memberikan solusi yang lebih efisien yang memungkinkan klasifikasi otomatis dan akurat menggunakan data numerik sederhana.

4.2. Data Understanding

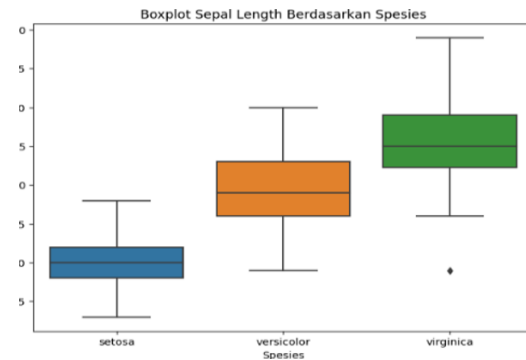
Berdasarkan data yang dikumpulkan dalam fase studi literatur untuk menentukan persyaratan sistem guna memecahkan masalah yang tepat, sumber data kemudian diproses dan dikembangkan secara menyeluruh menggunakan teknik yang dipilih oleh penulis untuk membantu dalam merancang sistem yang diidentifikasi. Fase ini melibatkan pengenalan data yang diperoleh dan mengidentifikasinya. Data tersebut berasal dari bunga iris dari 3 spesies yaitu setosa, versicolor, dan virginica.

4.3. Data Preparation

Data yang digunakan dalam penelitian ini adalah dataset bunga Iris yang setiap sampelnya memiliki empat fitur utama yaitu panjang daun kelopak, lebar daun kelopak, panjang petal, dan lebar petal. Pada fase ini dilakukan pembagian data menjadi data uji dan data latih untuk menguji akurasi model setelah pelatihan. Selanjutnya dilakukan preprocessing data seperti normalisasi, penanganan data yang hilang atau null, dan pelabelan. Langkah selanjutnya adalah memilih nilai K (jumlah tetangga terdekat yang dipertimbangkan dalam klasifikasi) merupakan langkah penting dalam pemodelan KNN. Model akan sensitif terhadap noise dan dapat mengalami *overfitting* apabila nilai K terlalu kecil. Jika nilai K terlalu besar, model dapat mengalami *underfitting* karena tidak sensitif terhadap pola data lokal.

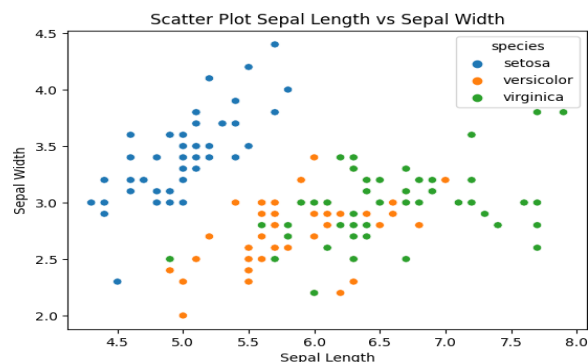
4.4. Modeling

Setelah data dipersiapkan dan nilai K dipilih, langkah berikutnya adalah melatih model KNN menggunakan training set. Pada tahap ini, algoritma KNN akan mempelajari pola dalam data dan menyimpan informasi tentang data training, sehingga dapat digunakan untuk mengklasifikasikan data baru. Hasil visualisasi model dapat dilihat pada gambar 2 dan 3.



Gambar 2. Hasil Boxplot Sepal

Gambar 2 merupakan boxplot yang menunjukkan distribusi panjang sepal (*sepal length*) pada tiga spesies bunga Iris, yakni setosa, versicolor, dan virginica. Dari visualisasi ini, terlihat bahwa Iris setosa memiliki panjang sepal yang lebih kecil dibandingkan dengan dua spesies lainnya, dengan rentang sekitar 4.5 hingga 5.8. Iris versicolor memiliki panjang sepal yang lebih besar dengan distribusi antara sekitar 5.0 hingga 7.0, sedangkan Iris virginica memiliki panjang sepal terbesar dengan rentang sekitar 5.5 hingga 8.0. Terdapat satu outlier pada spesies virginica yang memiliki panjang sepal lebih kecil dari mayoritas datanya. Boxplot ini membantu dalam memahami pola distribusi data dan variabilitas antar spesies sebelum diterapkan dalam model klasifikasi.

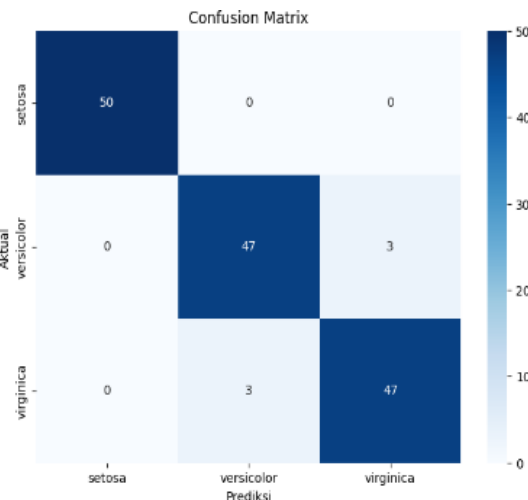


Gambar 3. Scatter Plot Sepal

Gambar 3 merupakan *scatter plot* yang menunjukkan hubungan antara panjang sepal (*Sepal Length*) dan lebar sepal (*Sepal Width*) pada tiga spesies bunga Iris: Setosa, Versicolor, dan Virginica. Terlihat bahwa Iris setosa (biru) memiliki distribusi yang cukup terpisah dari dua spesies lainnya, dengan sepal yang lebih pendek tetapi lebih lebar. Sementara itu, Iris versicolor (oranye) dan Iris virginica (hijau) memiliki distribusi yang lebih berdekatan, menunjukkan adanya tumpang tindih antara keduanya. Visualisasi ini membantu memahami pola distribusi data dan sejauh mana fitur panjang dan lebar sepal dapat digunakan untuk membedakan ketiga spesies.

4.5. Evaluation

Setelah tahap *modeling*, langkah selanjutnya yaitu berdasarkan data uji, model akan dievaluasi. Kinerja algoritma dapat diukur dari seberapa baik model mampu mengklasifikasikan data yang tidak pernah dilihat sebelumnya. Untuk mengevaluasi model, peneliti menggunakan *confusion matrix* untuk membandingkan hasil prediksi dengan label faktanya.



Gambar 4. Confusion Matrix

Gambar 4 di atas menunjukkan *confusion matrix* hasil klasifikasi tiga spesies bunga Iris menggunakan algoritma *K-Nearest Neighbor* (KNN). Matriks menunjukkan bahwa model memiliki kinerja yang kurat, dengan 50 sampel diklasifikasikan dengan benar. Untuk Iris versicolor, terdapat 47 sampel yang diklasifikasikan dengan benar, namun 3 sampel salah diklasifikasikan sebagai Iris virginica. Hal serupa terjadi pada Iris virginica, di mana 47 sampel diklasifikasikan dengan benar, tetapi 3 sampel salah diklasifikasikan sebagai Iris versicolor. Secara keseluruhan, model menunjukkan akurasi tinggi, dengan sedikit kesalahan dalam membedakan antara Iris versicolor dan Iris virginica, yang mungkin disebabkan oleh kemiripan fitur morfologi kedua spesies tersebut.

4.6. Deployment

Pada tahap ini dilakukan implementasi model ke dalam sebuah sistem. Berikut hasil dari pemodelan dan pengujian menggunakan algoritma KNN yang dapat dilihat pada gambar 5 sampai 7.

Gambar 5 merupakan tampilan antarmuka website untuk klasifikasi bunga Iris yang menggunakan algoritma *K-Nearest Neighbor* (KNN). Website ini menyediakan fitur utama seperti klasifikasi berdasarkan input fitur morfologi bunga, informasi tentang algoritma KNN, serta hasil klasifikasi yang ditampilkan secara interaktif. Untuk mendapatkan hasil klasifikasi spesies bunga Iris yang akurat, pengguna dapat memasukkan panjang sepal, lebar sepal, panjang petal, dan lebar petal, lalu memilih jumlah tetangga (K).

Confusion Matrix :

Aktual \ Prediksi	setosa	versicolor	virginica
setosa	50	0	0
versicolor	0	47	3
virginica	0	3	47

True Positive, True Negative, False Positive, False Negative :

Genus	True Positive	True Negative	False Positive	False Negative
Setosa	50	100	0	0
Virginica	47	100	3	3
Versicolor	47	100	3	3

Akurasi :

Keterangan	Nilai
Akurasi	96%
Diprediksi Benar	144
Total Data	150

Kesimpulan:

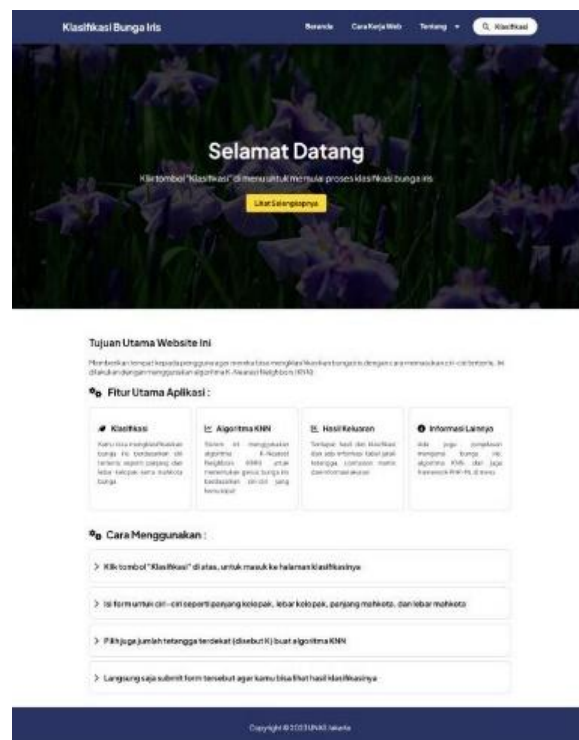
Berdasarkan penggunaan algoritma *K-Nearest Neighbors* (KNN) pada dataset Bunga Iris dan pengujian klasifikasi di atas, dapat disimpulkan bahwa dengan menggunakan nilai $K = 3$, telah ditentukan jenis bunga tersebut yaitu setosa, dengan tingkat ketepatan akurasi sebesar 96%, dengan fitur yang telah diinputkan sebagai berikut:

- Panjang Kelopak Bunga (Sepal Length): 4.6
- Lebar Kelopak Bunga (Sepal Width): 3.4
- Panjang Mahkota Bunga (Petal Length): 1.4
- Lebar Mahkota Bunga (Petal Width): 0.2

Hasil dari tabel jarak Euclidean menunjukkan 3 data terdekat dengan mengukur jarak antara fitur yang ada pada dataset dan fitur yang dimasukkan oleh pengguna. Maka, berdasarkan jarak tersebut, algoritma KNN telah berhasil mengklasifikasikan jenis bunga Iris yang paling sesuai:

No	Sample	Jarak	Genus
1	4.6 3.4 1.4 0.3	0.1	setosa
2	4.6 3.2 1.4 0.2	0.2	setosa
3	4.7 3.2 1.3 0.2	0.24494897427832	setosa

Gambar 5. Halaman Beranda



Gambar 6. Halaman Klasifikasi

Gambar 6 menunjukkan halaman hasil klasifikasi bunga Iris menggunakan algoritma *K-Nearest Neighbor* (KNN). Berdasarkan input fitur morfologi bunga, sistem mengklasifikasikan spesies bunga sebagai Iris setosa. Selain itu, tabel jarak Euclidean ditampilkan untuk menunjukkan perhitungan jarak antara sampel input dengan data dalam dataset, yang digunakan untuk menentukan klasifikasi akhir.

Hasil Klasifikasi Bunga Iris Menggunakan Algoritma KNN

Panjang Kelopak Bunga (Sepal Panjang): 4.5

Lebar Kelopak Bunga (Sepal Lebar): 3.4

Panjang Mahkota Bunga (Petal Panjang): 1.4

Lebar Mahkota Bunga (Petal Lebar): 0.2

Nilai K (KNN): 3

Genus Bunga : setosa

Hasil Klasifikasi

Table Jarak Euclidean :

No	Sample	Jarak	Perhitungan Jarak Euclidean	Genus
1	4.3, 3.1, 1.0, 0.2	0.1	$\sqrt{(4.3-4.5)^2 + (3.1-3.4)^2 + (1.0-1.4)^2 + (0.2-0.2)^2}$	setosa
2	4.6, 3.2, 1.4, 0.2	0.2	$\sqrt{(4.6-4.5)^2 + (3.2-3.4)^2 + (1.4-1.4)^2 + (0.2-0.2)^2}$	setosa
3	4.7, 3.1, 1.0, 0.2	0.24494897427832	$\sqrt{(4.7-4.5)^2 + (3.1-3.4)^2 + (1.0-1.4)^2 + (0.2-0.2)^2}$	setosa
4	4.8, 3.4, 1.4, 0.2	0.28242953647462	$\sqrt{(4.8-4.5)^2 + (3.4-3.4)^2 + (1.4-1.4)^2 + (0.2-0.2)^2}$	setosa
5	4.4, 3.2, 1.3, 0.2	0.3	$\sqrt{(4.4-4.5)^2 + (3.2-3.4)^2 + (1.3-1.4)^2 + (0.2-0.2)^2}$	setosa
6	4.7, 3.2, 1.4, 0.2	0.3	$\sqrt{(4.7-4.5)^2 + (3.2-3.4)^2 + (1.4-1.4)^2 + (0.2-0.2)^2}$	setosa
7	4.4, 3.1, 1.0, 0.2	0.2762276605484	$\sqrt{(4.4-4.5)^2 + (3.1-3.4)^2 + (1.0-1.4)^2 + (0.2-0.2)^2}$	setosa
8	4.8, 3.1, 1.4, 0.2	0.4023056258077	$\sqrt{(4.8-4.5)^2 + (3.1-3.4)^2 + (1.4-1.4)^2 + (0.2-0.2)^2}$	setosa
9	5.2, 4.1, 1.0, 0.2	0.4023056258077	$\sqrt{(5.2-4.5)^2 + (4.1-3.4)^2 + (1.0-1.4)^2 + (0.2-0.2)^2}$	setosa
10	5.2, 3.1, 1.4, 0.2	0.4023056258077	$\sqrt{(5.2-4.5)^2 + (3.1-3.4)^2 + (1.4-1.4)^2 + (0.2-0.2)^2}$	setosa

Gambar 7. Halaman Pengujian

Gambar 7 merupakan halaman pengujian untuk menampilkan hasil evaluasi klasifikasi bunga Iris disertai dengan kesimpulan.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa algoritma K-Nearest Neighbor (K-NN) dalam mengklasifikasikan tiga spesies bunga Iris dengan akurasi sebesar 96%. Model ini mampu mengidentifikasi Iris setosa dengan sangat baik, sementara terdapat sedikit kesalahan klasifikasi antara Iris versicolor dan Iris virginica, yang kemungkinan disebabkan oleh kemiripan fitur morfologi kedua spesies tersebut. Evaluasi menggunakan confusion matrix menunjukkan bahwa sebagian besar prediksi sesuai dengan label sebenarnya, mengindikasikan bahwa pemilihan nilai K yang optimal berpengaruh signifikan terhadap performa model. Selain itu, metode perhitungan jarak Euclidean membantu dalam menentukan kedekatan antar data, sehingga meningkatkan keakuratan klasifikasi.

Oleh karena itu, KNN dapat menjadi solusi yang sederhana namun efektif dalam pengklasifikasian bunga berdasarkan fitur morfologinya untuk meningkatkan akurasi model, eksplorasi nilai K dan penerapan teknik *cross-validation* dapat dilakukan guna memperoleh parameter terbaik. Selain itu, kombinasi KNN dengan algoritma lain seperti SVM atau *Random Forest* serta penggunaan feature engineering berbasis analisis citra dapat meningkatkan performa klasifikasi. Penelitian selanjutnya dapat menerapkan model yang sama namun dengan jumlah dataset yang berbeda.

DAFTAR PUSTAKA

- [1] L. Farokhah, "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," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1129–1136, 2020, doi: 10.25126/jtiik.202072608.
- [2] R. Sutri and A. Ramadhanu, "KLASIFIKASI BUAH KELAPA MUDA, KELAPA TUA, DAN BUAH," vol. 4307, no. 1, pp. 644–649, 2025.
- [3] A. P. Alkhairi, P., & Windarto, "Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019.
- [4] S. Santiastry et al, "Penerapan Algoritma Naive Bayes dan Metode CRISP-DM," vol. 3, pp. 10432–10439, 2024.
- [5] A. S. Roihan, A., Sunarya, P. A., & Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang," pp. 75–82, 2020.
- [6] W. Yahya & Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *J. Inform. dan Teknol.*, vol. 3, pp. 104–114, 2020.
- [7] J. Ddos and H. W. Aripadono, "JTIM: Jurnal Teknologi Informasi dan Multimedia Perbandingan Support Vector Machine, Random Forest Classifier, dan K-Nearest Neighbour dalam Pendeteksian Anomali," vol. 7, no. 1, pp. 23–33, 2025.
- [8] Y. A. Singgalen, "Analisis Performa Algoritma NBC, DT, SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1634–1646, 2022, doi: 10.47065/bits.v4i3.2766.
- [9] K. Ariyanti, D., & Iswardani, "Teks Mining untuk Klasifikasi Keluhan Masyarakat Menggunakan Algoritma Naive Bayes," *J. Komput. dan Inform.*, vol. 4, no. 3, pp. 125–132, 2020.
- [10] W. L. Y. Putro, H. F., Vlandari, R. T., & Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *Jurnal Teknol. Inf. dan Komun.*, vol. 8, no. 2, pp. 19–24, 2020, doi: https://doi.org/10.30646/tikomsin.v8i2.500.
- [11] Al-Arrafi, M. I., & Ramadhanu, A. (2025). Implementasi Metode Algoritma Principal Component Analysis (PCA) dan Algoritma K-Nearest Neighbor (KNN) dalam Klasifikasi Buah Jambu Madu Jambu Merah dan Manggis. *Journal of Science And Social Research*, 8(1), 817-822.