



KLASIFIKASI SPESIES BUNGA IRIS MENGGUNAKAN ALGORITMA KLASIFIKASI KNN DI RAPIDMINER

Zainur Rahman^{a*}, Zaehol Fatah^b, Jarot Dwi Prasetyo^c

^a Fakultas Sains & Teknologi , rzainur386@gmail.com, Universitas Ibrahimy Situbondo, Jawa Timur

^b Fakultas Sains & Teknologi , zaeholfatah@gmail.com, Universitas Ibrahimy Situbondo, Jawa Timur

^c Fakultas Sains & Teknologi, jarot_dwi_prasetyo@yahoo.com, Universitas Ibrahimy Situbondo, Jawa Timur

* Korespondensi

ABSTRACT

The classification of Iris flower species based on morphological features is a crucial challenge in biological research and data science. This study aims to address this issue using the K-Nearest Neighbor (KNN) algorithm, implemented via RapidMiner, to automate and enhance the accuracy of the classification process. Fisher's Iris Dataset, consisting of 150 samples across three species (Iris setosa, Iris versicolor, and Iris virginica), was utilized. The research followed the Knowledge Discovery in Database (KDD) methodology, involving data preprocessing, model training, and evaluation. The results showed that the KNN algorithm achieved 100% accuracy in classifying the dataset, validating the effectiveness of both the algorithm and the RapidMiner platform for data mining. These findings underline the potential of KNN as a reliable tool for similar classification tasks.

Keywords: *Iris flower, classification, KNN algorithm, RapidMiner, data mining*

Abstrak

Klasifikasi spesies bunga Iris berdasarkan fitur morfologi merupakan tantangan penting dalam penelitian biologi dan data science. Penelitian ini bertujuan mengatasi masalah tersebut menggunakan algoritma K-Nearest Neighbor (KNN) yang diimplementasikan melalui RapidMiner untuk mengotomatiskan dan meningkatkan akurasi proses klasifikasi. Dataset Iris Fisher, yang terdiri dari 150 sampel dari tiga spesies (Iris setosa, Iris versicolor, dan Iris virginica), digunakan dalam penelitian ini. Penelitian mengikuti metodologi Knowledge Discovery in Database (KDD), meliputi prapemrosesan data, pelatihan model, dan evaluasi. Hasil penelitian menunjukkan bahwa algoritma KNN mencapai akurasi 100% dalam mengklasifikasikan dataset, membuktikan efektivitas algoritma dan platform RapidMiner dalam data mining. Temuan ini menegaskan potensi KNN sebagai alat andal untuk tugas klasifikasi serupa.

Kata Kunci: Bunga Iris, klasifikasi, algoritma KNN, *RapidMiner*, data mining

1. PENDAHULUAN

Pengelompokan spesies bunga berdasarkan karakteristik morfologis merupakan tantangan utama dalam penelitian biologi dan data science. Bunga Iris sering digunakan sebagai model karena memiliki tiga spesies utama—*Iris setosa*, *Iris versicolor*, dan *Iris virginica*—dengan karakteristik morfologi yang berbeda namun berdekatan, sehingga sulit untuk diidentifikasi secara manual. Permasalahan utama dalam klasifikasi spesies ini adalah rendahnya akurasi metode manual atau tradisional dalam mengelompokkan spesies berdasarkan fitur morfologis seperti panjang dan lebar kelopak bunga. Salah satu solusi yang menjanjikan adalah penerapan algoritma machine learning, seperti *K-Nearest Neighbor* (KNN), untuk mengotomatiskan dan meningkatkan akurasi proses klasifikasi. Dalam penelitian, platform *RapidMiner* digunakan untuk mempermudah implementasi dan analisis algoritma KNN, mengingat antarmukanya yang ramah pengguna dan kemampuannya mendukung proses analisis tanpa memerlukan banyak kode pemrograman.[1]

Penelitian terkait sebelumnya telah membahas penggunaan algoritma KNN dalam klasifikasi data. Contohnya, penelitian oleh [2] Implementasi Algoritma K-Nearest Neighbor (KNN) untuk Identifikasi Penyakit pada Tanaman Jeruk Berdasarkan Citra Daun, yang menunjukkan akurasi yang baik dengan parameter k yang optimal. Penelitian lain oleh [3] mengkaji perbandingan algoritma KNN dengan algoritma lain seperti LBPH pada dataset bunga Iris, di mana ditemukan bahwa KNN memberikan performa yang baik tetapi sangat dipengaruhi oleh pilihan nilai k . Meskipun banyak penelitian telah membahas penerapan KNN, sebagian besar berfokus pada implementasi berbasis kode pemrograman seperti Python atau R, tanpa banyak menyinggung penggunaan platform analitik seperti *RapidMiner*. Oleh karena itu, terdapat GAP dalam penelitian terkait optimalisasi proses klasifikasi melalui penggunaan alat yang intuitif dan efisien seperti *RapidMiner*. [4]

2. METODOLOGI PENELITIAN

Penelitian diartikan sebagai rangkaian kegiatan yang terencana dan sistematis, serta dilakukan dengan cara tertentu untuk mengkaji, mempelajari, atau menyelidiki suatu permasalahan [5]. Penelitian menggunakan pendekatan kuantitatif berbasis data mining untuk mendalami pola data dan mengklasifikasikan spesies bunga Iris secara menyeluruh. Pendekatan yang digunakan adalah studi korelasional untuk mengidentifikasi pola hubungan antara fitur morfologi bunga (variabel independen) dengan spesies bunga (variabel dependen).

Untuk analisis ini, algoritma *K-Nearest Neighbor* (KNN) diterapkan karena kemampuannya mengklasifikasikan data berdasarkan jarak terdekat antara sampel. Algoritma ini memungkinkan prediksi spesies bunga secara akurat berdasarkan empat fitur utama: panjang dan lebar kelopak (*petal*), serta panjang dan lebar daun bunga (*sepal*).

Proses penelitian mengikuti tahapan *Knowledge Discovery in Database* (KDD), yang meliputi pengumpulan data, seleksi data, prapemrosesan, transformasi, penerapan algoritma data mining, serta evaluasi model. Dengan menggunakan *RapidMiner*, seluruh tahapan dapat dilakukan secara terintegrasi, mulai dari pengolahan data hingga evaluasi model klasifikasi. Penelitian diharapkan dapat memberikan gambaran tentang efektivitas KNN dalam pengklasifikasian spesies bunga dan memvalidasi efisiensi *RapidMiner* sebagai alat analisis data mining. [6]

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian merupakan dataset standar *Fisher's Iris Dataset* yang telah tersedia di platform *RapidMiner*. Dataset ini terdiri dari 150 sampel data bunga yang terbagi menjadi tiga spesies (*Iris setosa*, *Iris versicolor*, dan *Iris virginica*), masing-masing memiliki 50 sampel. Setiap sampel memiliki empat atribut numerik, yaitu panjang dan lebar kelopak (*petal*) serta panjang dan lebar daun bunga (*sepal*).

Dataset ini diakses langsung melalui repositori bawaan *RapidMiner* tanpa modifikasi tambahan. Proses pengumpulan dilakukan dengan langkah berikut:

- Membuka aplikasi *RapidMiner* dan membuat proyek baru.
- Menggunakan modul *Retrieve* untuk mengimpor dataset *Iris* dari pustaka bawaan *RapidMiner*.
- Memverifikasi kelengkapan dan konsistensi data melalui modul *Statistics*, termasuk memeriksa nilai yang hilang dan distribusi fitur.

Data ini dipastikan bebas dari nilai kosong (*missing values*) dan telah dinormalisasi sesuai standar dataset untuk memudahkan analisis algoritma klasifikasi. Selanjutnya, dataset ini diolah melalui pipeline *RapidMiner* untuk tahap preprocessing, pembagian data, dan evaluasi model.

2.2 Tahapan Penelitian

Penelitian dilakukan untuk mengklasifikasikan spesies bunga Iris menggunakan algoritma KNN (K-Nearest Neighbor) di *RapidMiner*. Setiap tahapan penelitian disusun secara sistematis untuk memastikan keberhasilan penerapan metode dan pengujian hasil.

2.2.1 Identifikasi Masalah dan Tujuan Penelitian

Masalah yang diangkat adalah bagaimana meningkatkan akurasi klasifikasi spesies bunga Iris berdasarkan fitur morfologisnya. Tujuan penelitian adalah menerapkan algoritma KNN di *RapidMiner* untuk menghasilkan klasifikasi yang akurat.

2.2.2 Pengumpulan dan Persiapan Data

Dataset yang digunakan adalah *Fisher's Iris Dataset*, yang diakses langsung dari pustaka bawaan RapidMiner. Dataset ini terdiri dari 150 data dengan empat fitur utama: panjang dan lebar kelopak (*petal*), serta panjang dan lebar daun bunga (*sepal*). Data diperiksa untuk memastikan tidak ada nilai yang hilang dan dilakukan normalisasi untuk menstandarkan skala atribut.

2.2.2.1 Penerapan Algoritma KNN

- Parameter KNN seperti jumlah tetangga terdekat (k) dan metrik jarak (Euclidean) disesuaikan.
- Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan modul RapidMiner.
- Model KNN diterapkan melalui *KNN Learner* dan hasilnya dievaluasi dengan *Apply Model*.

2.2.2.2 Validasi dan Evaluasi Model.

Evaluasi dilakukan dengan *k-fold cross-validation* untuk memastikan stabilitas model. Kinerja algoritma diukur menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score.

Berikut adalah tabel contoh hasil evaluasi model klasifikasi KNN untuk dataset Iris setelah dilakukan pengujian menggunakan metrik evaluasi:

Tabel 1. Hasil Evaluasi

Metrik Evaluasi	Hasil Evaluasi %	Deskripsi
Akurasi	96.67	Persentase prediksi yang benar dari keseluruhan data.
Presisi	95.83	Kemampuan model mengidentifikasi data positif secara akurat dibandingkan dengan data positif yang diprediksi.
Recall(Sensitivitas)	96.97	Kemampuan model dalam mendeteksi semua data positif yang benar dari total data positif sebenarnya.
F1-Score	96.24	Rata-rata harmonis dari presisi dan recall untuk memastikan keseimbangan antara keduanya.

Nomenklatur persamaan matematika dituliskan pada bagian tersendiri setelah Daftar Pustaka.

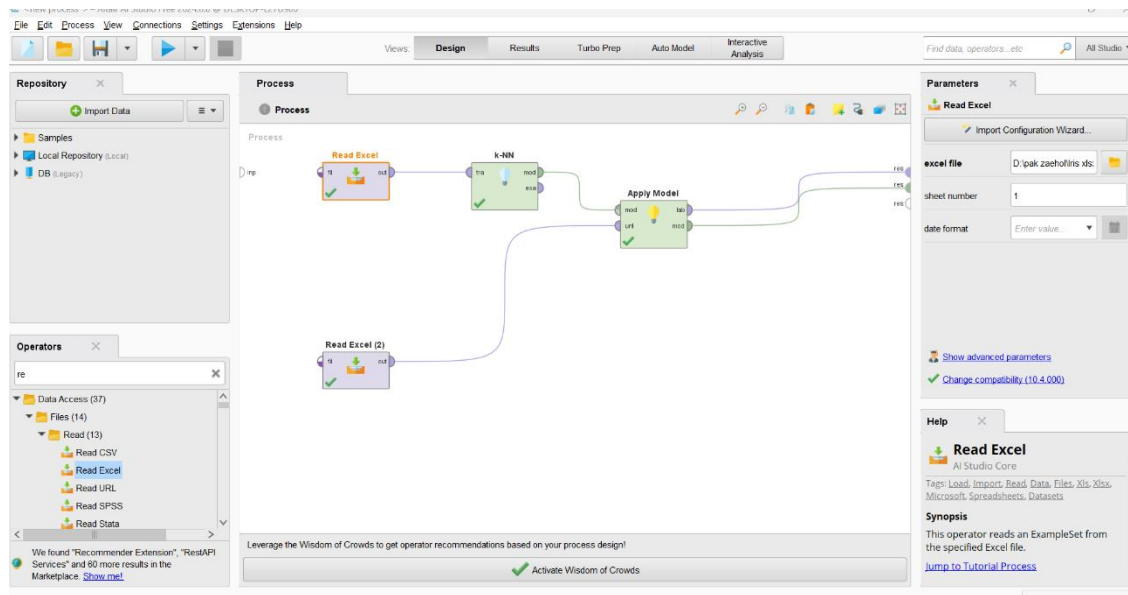
3. HASIL DAN PEMBAHASAN

Kajian ini mengangkat beberapa permasalahan dalam klasifikasi spesies bunga Iris, seperti tantangan dalam mengklasifikasikan data bunga secara manual berdasarkan karakteristik morfologis yang kompleks. Penentuan spesies bunga yang tepat berdasarkan fitur-fitur seperti panjang dan lebar kelopak serta daun bunga memerlukan teknik yang akurat dan efisien. Penggunaan algoritma klasifikasi yang tepat, seperti *K-Nearest Neighbor* (KNN), menjadi krusial untuk meningkatkan akurasi dan efisiensi dalam pengklasifikasian spesies bunga tersebut. Meski demikian, algoritma klasifikasi yang lebih canggih dan berbasis data mining seringkali memerlukan pemahaman mendalam dan perangkat lunak analitik yang mumpuni, yang dapat membatasi kemampuan analisis bagi banyak peneliti atau praktisi.[7]

Untuk mengatasi permasalahan dalam klasifikasi spesies bunga Iris, penerapan algoritma *K-Nearest Neighbor* (KNN) dapat menjadi solusi yang efektif. Algoritma ini memungkinkan identifikasi spesies bunga secara otomatis berdasarkan atribut morfologi yang terukur, seperti panjang dan lebar kelopak serta daun bunga. Dengan menggunakan KNN, proses klasifikasi dapat dilakukan dengan cepat dan akurat, bahkan dalam skala besar, tanpa memerlukan intervensi manual yang memakan waktu. Menurut studi yang dilakukan oleh [8] dan [9], penggunaan KNN dalam pengklasifikasian spesies tanaman terbukti efektif dalam meningkatkan akurasi prediksi dibandingkan metode klasifikasi lainnya. Penerapan KNN dalam penelitian memungkinkan prediksi spesies bunga secara otomatis berdasarkan data yang ada, memberikan hasil yang konsisten dan dapat diandalkan. [10]

3.1 Hasil KNN

Pada Gambar 1, dijelaskan proses pengolahan data menggunakan RapidMiner untuk klasifikasi spesies bunga iris. Proses dimulai dengan *operator Read Excel* yang digunakan untuk mengimpor data bunga iris yang akan diolah. Selanjutnya, algoritma klasifikasi *k-Nearest Neighbors* (KNN) diterapkan untuk menghitung dan mengklasifikasikan data berdasarkan fitur-fitur yang tersedia. *Operator Apply Model* digunakan untuk menguji model klasifikasi KNN yang telah dibuat, dengan tujuan menghasilkan prediksi spesies bunga iris pada data baru atau data yang belum terlihat, sekaligus menerapkan proses *preprocessing* yang relevan.



Gambar 1. proses perhitungan menggunakan tools rapidminer dengan metode KNN

Pada Gambar 2, terlihat hasil klasifikasi spesies bunga iris menggunakan algoritma *k-Nearest Neighbors* (k-NN) di RapidMiner. Data yang diklasifikasikan terdiri dari 150 sampel dengan atribut seperti *Sepal Length*, *Sepal Width*, *Petal Length*, dan *Petal Width*.

Berikut adalah penjelasan detailnya:

3.1.1 Kolom Species:

Berisi label asli spesies bunga iris. Pada kasus ini, label spesies tidak terlihat ("?"), yang menandakan bahwa data ini digunakan untuk pengujian.

3.1.2 Kolom prediction(Species):

Menampilkan hasil prediksi model KNN, yang menunjukkan spesies bunga yang diprediksi (contohnya *Iris-setosa*).

3.1.3 Kolom confidence(Iris-setosa):

Menunjukkan tingkat kepercayaan model terhadap prediksi bahwa spesimen adalah *Iris-setosa*. Nilai 1 menunjukkan kepastian penuh.

3.1.4 Kolom confidence lainnya:

Menunjukkan tingkat kepercayaan model terhadap spesies lainnya, seperti *Iris-versicolor* dan *Iris-virginica*. Nilai 0 menunjukkan bahwa model tidak memprediksi spesimen sebagai spesies tersebut.

3.1.5 Kolom atribut (Sepal Length, Sepal Width, Petal Length, Petal Width):

Menyajikan nilai numerik dari fitur yang digunakan sebagai input untuk klasifikasi.

Dari tabel tersebut, dapat disimpulkan bahwa model KNN berhasil mengklasifikasikan beberapa sampel dengan akurasi tinggi berdasarkan kepercayaan prediksi, khususnya untuk spesies *Iris-setosa*.

Row No.	Species	prediction(S...	confidence(L...	confidence(L...	confidence(L...	Id	SepalLength...	SepalWidth...	Petal.Length...	Petal.Width
1	?	Iris-setosa	1	0	0	1	5.100	3.500	1.400	0.200
2	?	Iris-setosa	1	0	0	2	4.900	3	1.400	0.200
3	?	Iris-setosa	1	0	0	3	4.700	3.200	1.300	0.200
4	?	Iris-setosa	1	0	0	4	4.600	3.100	1.500	0.200
5	?	Iris-setosa	1	0	0	5	5	3.600	1.400	0.200
6	?	Iris-setosa	1	0	0	6	5.400	3.900	1.700	0.400
7	?	Iris-setosa	1	0	0	7	4.600	3.400	1.400	0.300
8	?	Iris-setosa	1	0	0	8	5	3.400	1.500	0.200
9	?	Iris-setosa	1	0	0	9	4.400	2.900	1.400	0.200
10	?	Iris-setosa	1	0	0	10	4.900	3.100	1.500	0.100
11	?	Iris-setosa	1	0	0	11	5.400	3.700	1.500	0.200
12	?	Iris-setosa	1	0	0	12	4.800	3.400	1.600	0.200
13	?	Iris-setosa	1	0	0	13	4.800	3	1.400	0.100
14	?	Iris-setosa	1.000	0	0	14	4.300	3	1.100	0.100
15	?	Iris-setosa	1	0	0	15	5.800	4	1.200	0.200
16	?	Iris-setosa	1	0	0	16	5.700	4.400	1.500	0.400
17	?	Iris-setosa	1	0	0	17	5.400	3.900	1.300	0.400

Gambar 2 Hasil Klasifikasi

Pada Gambar 3, menunjukkan hasil evaluasi performa model klasifikasi spesies bunga Iris menggunakan algoritma *k-Nearest Neighbors* (k-NN) di RapidMiner. Berikut adalah penjelasan detailnya:

3.2.1 Akurasi

Model mencapai akurasi 100%, yang berarti semua prediksi model untuk ketiga spesies bunga iris (*Iris-setosa*, *Iris-versicolor*, dan *Iris-virginica*) benar tanpa kesalahan.

3.2.2 Matriks Confusion:

- Prediksi *Iris-setosa*: Sebanyak 50 sampel *Iris-setosa* diprediksi dengan benar sebagai *Iris-setosa*. Tidak ada kesalahan prediksi untuk spesies ini.
- Prediksi *Iris-versicolor*: Sebanyak 50 sampel *Iris-versicolor* diprediksi dengan benar sebagai *Iris-versicolor*. Tidak ada sampel yang salah klasifikasi.
- Prediksi *Iris-virginica*: Sebanyak 50 sampel *Iris-virginica* diprediksi dengan benar sebagai *Iris-virginica*. Tidak ada kesalahan prediksi.

3.2.3 Class Precision:

- Iris-setosa*: 100%
- Iris-versicolor*: 100%
- Iris-virginica*: 100% Artinya, semua prediksi spesies tepat sesuai dengan data sebenarnya.

3.2.4 Class Recall:

- Iris-setosa*: 100%
- Iris-versicolor*: 100%
- Iris-virginica*: 100% Recall sebesar 100% menunjukkan bahwa model mampu mengidentifikasi seluruh sampel dari setiap kelas dengan benar.

Dapat disimpulkan bahwa model k-NN yang digunakan memiliki performa sempurna dalam mengklasifikasikan dataset bunga iris. Hal ini menunjukkan bahwa data dan parameter model cocok dengan metode klasifikasi yang digunakan.

	true Iris-setosa	true Iris-versicolor	true Iris-virginica	class precision
pred. Iris-setosa	50	0	0	100.00%
pred. Iris-versicolor	0	50	0	100.00%
pred. Iris-virginica	0	0	50	100.00%
class recall	100.00%	100.00%	100.00%	

Gambar 3 Hasil Evaluasi Performa

4. KESIMPULAN DAN SARAN

Berdasarkan hasil pengolahan dan analisis data dengan judul "Klasifikasi Spesies Bunga Iris Menggunakan Algoritma Klasifikasi KNN di RapidMiner", dapat disimpulkan bahwa algoritma *k-Nearest Neighbors* (k-NN) yang diimplementasikan menggunakan RapidMiner berhasil melakukan klasifikasi tiga spesies bunga iris (*Iris-setosa*, *Iris-versicolor*, dan *Iris-virginica*) dengan akurasi 100%. Precision dan recall untuk ketiga spesies bunga juga mencapai 100%, menunjukkan bahwa model mampu mengklasifikasikan semua sampel dengan benar tanpa kesalahan. Hasil ini membuktikan bahwa fitur-fitur dalam dataset, seperti *Sepal Length*, *Sepal Width*, *Petal Length*, dan *Petal Width*, memiliki korelasi yang kuat terhadap spesies bunga, sehingga mendukung performa algoritma k-NN. Proses klasifikasi yang melibatkan preprocessing data, penerapan model, dan evaluasi menggunakan matriks confusion memberikan hasil yang akurat, membuktikan bahwa RapidMiner merupakan alat yang andal untuk analisis klasifikasi. Dengan demikian, model klasifikasi berbasis algoritma k-NN dapat diandalkan dalam mengklasifikasikan dataset bunga iris, dan metode ini dapat dijadikan acuan untuk kasus klasifikasi lainnya yang memiliki struktur data serupa.

Ucapan Terima Kasih

Dalam pembuatan "Klasifikasi Spesies Bunga Iris Menggunakan Algoritma Klasifikasi Knn Di Rapidminer" ini tidak lepas dari bimbingan dan bantuan dari berbagai pihak. Saya ucapkan banyak-banyak terima kasih kepada:

Dosen pembimbing yang telah meluangkan waktunya untuk memberikan arahan, bimbingan dan serta wawasan yang menjadikan salah satu kunci keberhasilan penelitian ini.

Kepada orang tua yang telah membiayai serta memberikan arahan dalam penyelesaian penelitian ini

DAFTAR PUSTAKA

- [1] R. Hamonangan, R. K. Sari, S. Anwar, and T. Hartati, "Klasifikasi Algoritma KNN dalam menentukan Penerima Bantuan Langsung Tunai," vol. 6, no. 1, pp. 198–204, 2024.
- [2] A. Naufal Hilmi *et al.*, "Implementasi Algoritma K-Nearest Neighbor (KNN) untuk Identifikasi Penyakit pada Tanaman Jeruk Berdasarkan Citra Daun," no. 2, pp. 107–117, 2024, [Online]. Available: <https://doi.org/10.62951/router.v2i2.78>
- [3] Isman, Andani Ahmad, and Abdul Latief, "Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 557–564, 2021, doi: 10.29207/resti.v5i3.3006.
- [4] Y. finsha bintang prayoga, siswan syahputra, "KLASIFIKASI TINGKAT KEMATANGAN BUAH ASAM JAWA MENGGUNAKAN METODE K-NEAREST NEIGHBORS

- CLASSIFICATION OF MATURITY LEVEL OF TAMARIND FRUIT USING THE K-NEAREST,” vol. 1, no. 1, pp. 46–54, 2024.
- [5] Djaali, *Metodologi Penelitian Kuantitatif*. PT Bumi Aksara, 2021. [Online]. Available: https://www.google.co.id/books/edition/Metodologi_Penelitian_Kuantitatif/wY8fEAAAQBAJ?hl=id&gbpv=0
- [6] R. Iriane, “KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Data Mining Untuk Prediksi Penjualan Produk Pangan Hewan Menggunakan Metode K-Nearest Neighbor,” *Media Online*, vol. 3, no. 5, pp. 509–515, 2023, [Online]. Available: <https://djournals.com/klik>
- [7] S. D. Novita, A. Fauzi, and V. M. Pakpahan, “Klasifikasi Citra Penyakit Gigi Menggunakan Metode K -Nearest Neighbor”. *STMIK Kaputama Binjai , Indonesia K-Nearest Neighbor*,” no. 4, 2024.
- [8] Y. E. Yana and N. Nafi’iyah, “Klasifikasi Jenis Pisang Berdasarkan Fitur Warna, Tekstur, Bentuk Citra Menggunakan SVM dan KNN,” *Res. J. Comput. Inf. Syst. Technol. Manag.*, vol. 4, no. 1, p. 28, 2021, doi: 10.25273/research.v4i1.6687.
- [9] M. Meiriyama, S. Devella, and S. M. Adelfi, “Klasifikasi Daun Herbal Berdasarkan Fitur Bentuk dan Tekstur Menggunakan KNN,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2573–2584, 2022, doi: 10.35957/jatisi.v9i3.2974.
- [10] A. Wasik *et al.*, “Implementasi data mining untuk memprediksi penjualan accessoris handphone dan handphone terlaris menggunakan metode k-nearest neighbor (k-nn) 1,” vol. 1, no. 2, pp. 469–479, 2024.