

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Penelitian Terdahulu**

Penelitian mengenai klasifikasi citra telah dilakukan oleh beberapa peneliti terdahulu, antara lain sebagai berikut:

Penelitian yang dilakukan oleh Feri Wibobwo, Agung Purwo Wicaksono, Lahan Adi Purwanto pada tahun 2021 dengan judul “Klasifikasi Tanaman Beringin (*Ficus Benjamina*) Berdasarkan Citra Daun Menggunakan Algoritma *K-Nearest Neighbors*”. Tujuan dari penelitian ini untuk mengklasifikasikan beringin putih berdasarkan bentuk dan teksturnya dan menentukan kedekatan antara data uji dengan data latih menggunakan metode *K-Nearest Neighbor*. Penelitian ini mendapatkan akurasi sebesar 56,25% (Wibowo et al., 2021).

Penelitian selanjutnya dilakukan oleh Ahmad Rifan Firdaus, Moch Lutfi, Muhammad Faishol Amrulloh pada tahun 2022 dengan judul “Tanaman Klengkeng Berdasarkan Ciri Tekstur Daun Menggunakan Metode *Adaptive Neuro Fuzzy Iverence System* (AFIS)”. Tujuan dari penelitian ini untuk mengklasifikasikan citra daun tanaman klengkeng berdasarkan jenis varietasnya dan menentukan kedekatan data uji dengan data latih menggunakan metode *Fuzzy*. Penelitian ini mendapatkan akurasi sebesar 79.5% (Amrulloh, 2022).

Penelitian selanjutnya dilakukan oleh Valian Yoga Pudya Ardhana, Joni Saputra, Mafriansyah pada tahun 2022 dengan judul “Klasifikasi Jenis Mangga Berdasarkan Tekstur Tulang Daun Me

nggunakan Metode *Learning Vektor Quantization* (LVQ)”. Tujuan dari penelitian ini untuk mengklasifikasikan jenis mangga berdasarkan tekstur tulang daun dan menentukan kedekatan data uji dengan data latih menggunakan metode *Learning Vektor Quantization*. Penelitian ini mendapatkan akurasi sebesar 100.80% (Ardhana et al., 2022).

Penelitian selanjutnya dilakukan oleh Andi Yulia Muniar pada tahun 2019 dengan judul “Klasifikasi Tanaman Jeruk Berdasarkan Fitur Tekstur Daun Menggunakan Metode *K-Nearest Neighbors*”. Tujuan dari penelitian ini untuk mengklasifikasikan tanaman jeruk berdasarkan fitur tekstur daun dan menentukan kedekatan data uji dengan data latih menggunakan metode *K-Nearest Neighbors*. Penelitian ini mendapatkan akurasi sebesar 81.48% (Muniar, 2019).

Penelitian selanjutnya dilakukan oleh Christhy Atika Sari, Eko Hari Rachmawanto pada tahun 2023 dengan judul “Klasifikasi Jenis Pepaya Berdasarkan Citra Daun Menggunakan Metode *Naive Bayes*”. Tujuan dari penelitian ini untuk mengklasifikasikan jenis buah pepaya berdasarkan citra daun dan menentukan kedekatan data uji data latih menggunakan metode *Naive Bayes*. Penelitian ini mendapatkan akurasi sebesar 96.93% (Sari & Rachmawanto, 2021).

Penelitian selanjutnya dilakukan oleh Ananta Dwi Prayoga Alwy, Bukhati Naufal Nur AG, Silvia Andriani, Fhatiah Adiba, Andi Baso Kaswar pada tahun 2022 dengan judul “Klasifikasi Kematangan Daun Selada Berdasarkan Fitur Warna Menggunakan metode *K-Nearest Neighbors*”. Tujuan dari penelitian ini untuk mengklasifikasikan kematangan daun selada berdasarkan fitur warna dan

menentukan kedekatan data uji dan data latih menggunakan metode *K-Nearest Neighbors*. Penelitian ini mendapatkan akurasi sebesar 95% (Alwy et al., 2022).

Penelitian selanjutnya dilakukan oleh Yusuf Eka Yana, Nur Nafi'iyah pada tahun 2021 dengan judul “Klasifikasi Jenis Pisang Berdasarkan Fitur Warna, Tekstur, Bentuk Citra Menggunakan metode SVM dan KNN”. Tujuan dari penelitian ini untuk mengklasifikasikan jenis-jenis pisang berdasarkan fitur warna, tekstur dan bentuk dan menentukan kedekatan data uji dan data latih menggunakan metode SVM dan KNN. Penelitian ini mendapatkan akurasi sebesar 45,24% (Yana & Nafi'iyah, 2021).

Penelitian selanjutnya dilakukan oleh Unga Hasna, Alda Cendekia Siregar, Barry Ceasar Octariadi pada tahun 2023 dengan judul “Klasifikasi Bentuk Daun Tanaman Begonia (*Begoniaceae*) Menggunakan Metode *K-Nearest Neighbors*”. Tujuan dari penelitian ini untuk mengklasifikasikan jenis tanaman begonia berdasarkan bentuk daun dan menentukan kedekatan data uji dan data latih menggunakan metode *K-Nearest Neighbors*. Penelitian ini mendapatkan akurasi 100% (Hasna et al., 2022).

Penelitian selanjutnya dilakukan oleh Meiriyama Siska Devella, Sandra Mareza Adelfi pada tahun 2022 dengan judul “Klasifikasi Daun Herbal Berdasarkan Fitur Bentuk Dan Tekstur Menggunakan metode KNN”. Tujuan dari penelitian ini untuk mengklasifikasikan daun tanaman herbal dan menentukan kedekatan data uji dan data latih menggunakan metode *K-Nearest Neighbors*. Penelitian ini mendapatkan akurasi 92,67% (Meiriyama et al., 2022).

Penelitian selanjutnya dilakukan oleh Bahtiar Adi Prasetya, Zilvanhisna Emka Fitri, Abdul Madjid, Arizal Mujibtamala Nanda Imron pada tahun 2022 dengan judul “Ensiklopedia Digital Varietas Ubi Jalar Berdasarkan Klasifikasi Citra Daun Menggunakan metode *K-Nearest Neighbors*”. Tujuan dari penelitian ini untuk membuat sebuah ensiklopedia digital varietas ubi jalar berdasarkan bentuk daun dan menentukan kedekatan data uji dan data latih menggunakan metode *K-Nearest Neighbors*. Penelitian ini mendapatkan akurasi 95% (Prasetya et al., 2022).

Penelitian selanjutnya dilakukan oleh Ninda Silvia Cahyani, Nurul Hidayat, Edy Santoso pada tahun 2023 dengan judul “Klasifikasi Penyakit Tanaman Kacang Tanah menggunakan metode *Modified K-Nearest Neighbors*”. Tujuan dari penelitian ini untuk klasifikasi penyakit pada tanaman kacang tanah dan menentukan kedekatan data uji dan data latih menggunakan metode *Modified K-Nearest Neighbor*. Penelitian ini mendapatkan akurasi 97% (Cahyani et al., 2023).

Berikut ini adalah rangkuman dari beberapa penelitian yang telah dijabarkan sebelumnya dan dapat di lihat pada tabel 2.1 berikut ini :

Tabel 2.1 Perbandingan Penelitian

No	Nama Penulis	Judul Penelitian	Metode	Hasil
1	Feri Wibowo, Agung Purwo Wicaksono, Lahan Adi Purwanto (2021).	Klasifikasi Tanaman Beringin ( <i>Ficus Benjamina</i> ) Berdasarkan Citra Daun Menggunakan Algoritma <i>K-Nearest Neighbors</i>	<i>K-Nearest Neighbors</i>	Berdasarkan hasil pengujian klasifikasi tanaman beringin menggunakan Algoritma <i>K-Nearest Neighbors</i> mendapatkan akurasi paling rendah yaitu 56,25%.
2	Ahmad Rifan Firdaus, Moch Lutfi, Muhamad Faishol Amrulloh (2022).	Tanaman Klengkeng Berdasarkan Ciri Tekstur Daun Menggunakan Metode <i>Adaptive Neuro Fuzzy Iverence System</i> (AFIS)	<i>Adaptive Neuro Fuzzy Iverence System</i> (AFIS)	Berdasarkan hasil penelitian ini menggunakan 80 data citra, kemampuan pengenalan dengan rata-rata sebesar 79,5%.
3	Valian Yoga Pudya Ardhana, Joni Saputra, M,	Klasifikasi Jenis Mangga Berdasarkan Tekstur Tulang	<i>Learning Vektor Quantization</i> (LVQ)	Berdasarkan hasil dari perhitungan yang didapatkan cluster yang tersisa

	Afriansyah (2022).	Daun Menggunakan Metode <i>Learning Vektor Quantization</i> (LVQ)		adalah cluster 3 dengan nilai centroid R =151,67 G = 145 dan B = 153,33
4	Andi Yulia Muniar (2019).	Klasifikasi Tanaman Jeruk Berdasarkan Fitur Tekstur Daun Menggunakan Metode <i>K-Nearest Neighbors</i>	<i>K-Nearest Neighbors</i>	Berdasarkan hasil penelitian dapat disimpulkan bahwa metode ekstraksi ciri fitur tekstur daun serta Algoritma <i>K-Nearest Neighbors</i> dapat digunakan untuk mengklasifikasikan tanaman jeruk menggunakan euclidean distance pada parameter k=1 dengan hasil akurasi tertinggi sebesar 81,48
5	Christhy Atika Sari, Eko Hari Rachmawanto (2023).	Klasifikasi Jenis Pepaya Berdasarkan Citra Daun Menggunakan Metode <i>Naive</i>	<i>Naive Bayes</i>	Berdasarkan hasil penelitian klasifikasi jenis pepaya berdasarkan citra daun menggunakan

		<i>Bayes</i>		Metode <i>Naive Bayes</i> dari data <i>testing, confusion matrix</i> dan akurasi yang sangat baik yaitu 96%.
6	Ananta Dwi Prayoga Alwy, Bukhari Naufal Nur AG, Silvia Adriani, Fhatiah Adiba, Andi Baso Kaswar(2022).	Klasifikasi Kematangan Daun Selada Berdasarkan Fitur Warna Menggunakan metode <i>K-Nearest Neighbors</i>	<i>K-Nearest Neighbors</i>	Dalam Penelitian ini kami menggunakan 60 dataset citra yang dibagi menjadi 40 data latih dan 20 data uji berdasarkan hasil uji dengan menggunakan metode <i>K-Nearest Neighbor</i> didapatkan akurasi pengujian sebesar 98 %.
7	Yusuf Eka Yana, Nur Nafi'iyah (2021	Klasifikasi Jenis Pisang Berdasarkan Fitur Warna, Tekstur, Bentuk Citra Menggunakan metode SVM dan <i>KNN</i>	<i>SVM dan KNN</i>	Berdasarkan hasil klasifikasi jenis pisang dengan algoritma <i>SVM</i> mendapatkan akurasi 58,33%.

8	Unga Hasna, Alda Cendekia Siregar, Barry Ceasar Oktariadi (2023).	Klasifikasi Bentuk Daun Tanaman Begonia ( <i>Begoniaceae</i> ) Menggunakan Metode <i>K-Nearest Neighbors</i>	<i>K-Nearest Neighbors</i>	Berdasarkan hasil penelitian klasifikasi bentuk daun tanaman begonia menggunakan Metode <i>K-Nearest Neighbors</i> mendapatkan Akurasi sebesar 100%.
9	Meiriyama, Siska Devella, Sandra Mareza Adelfi (2022).	Klasifikasi Daun Herbal Berdasarkan Fitur Bentuk Dan Tekstur Menggunakan metode KNN	KNN	Berdasarkan implementasi pengujian yang dilakukan, dapat disimpulkan bahwa sistem dapat mengklasifikasikan jenis daun tanaman herbal dengan baik. Berdasarkan hasil pengujian, HOG + KNN yang memperoleh akurasi 0,93% LBP+KNN yang memperoleh akurasi 0,89%.
10	Bahtiar Adi Prasetya,	Ensiklopedia Digital Varietas	<i>K-Nearest Neighbors</i>	Berdasarkan hasil penelitian yang di

	Zilvanhisna Emka Ftri, Abdul Madjid, Arizal Mujibtamala Nanda Imron (2022).	Ubi Jalar Berdasarkan Klasifikasi Citra Daun Menggunakan metode <i>K-Nearest Neighbors</i>		peroleh Metode <i>k-Nearest Neighbors</i> mampu mengklasifikasikan citra daun ubi jalar untuk ensiklopedia digital dengan tingkat akurasi 95%.
11	Ninda Silvia Tri Cahyani, Nurul Hidayat, Edy Santoso (2023).	Klasifikasi Penyakit Tanaman Kacang Tanah menggunakan metode <i>Modified K-Nearest Neighbors</i>	<i>Modified K-Nearest Neighbors</i>	Berdasarkan hasil penelitian Klasifikasi Tanaman Kacang Tanah menggunakan metode <i>Modified K-Nearest Neighbors</i> nilai K mendapatkan akurasi 96,67%

Pada penelitian yang dilakukan tentang Klasifikasi Tanaman Beringin Berdasarkan Citra Daun Menggunakan Algoritma *K-Nearest Neighbors* mengacu pada penelitian sebelumnya yang dilakukan oleh (Wibowo et al.,2021) dengan judul Klasifikasi tanaman beringin (*Ficus Benjamina*) berdasarkan citra daun menggunakan algoritma *k-nearest neighbors*. Hasil dari penelitian ini adalah dengan mendapatkan akurasi paling rendah yaitu 56,25%. Maka perbedaan penelitian saat ini dengan penelitian rujukan terdapat pada perbedaan tingkat

akurasi, dimana hasil penelitian saat ini mendapatkan akurasi yang paling rendah yaitu terdapat pada *number of folds* 2 dan 5 dengan pengklasifikasian daun beringin putih mendapatkan akurasi sebesar 82,5%.

## 2.2 Algoritma *K-Nearest Neighbors*

Algoritma *K-Nearest Neighbors* adalah salah satu algoritma klasifikasi dalam *machine learning* yang termasuk dalam kelompok *supervised learning*. KNN bekerja berdasarkan prinsip bahwa objek yang serupa cenderung berada dalam kelas yang sama. Dalam KNN, klasifikasi data uji dilakukan dengan mencari nilai *K* tetangga terdekat dari data uji dalam ruang fitur dan menentukan kelas mayoritas dari *K* tetangga tersebut. Rumus-rumus dalam algoritma KNN mencakup perhitungan jarak antara data uji dengan data latih, serta tahapan-tahapan lain yang terlibat dalam proses klasifikasi antara lain (Yusuf et al., 2023)

1. Cara kerja *K-Nearest Neighbors* :
  - a. Pilih nomor *K-Nearest Neighbors* atau terdekat.
  - b. Hitung jarak *Euclidean* dari jumlah *K* terdekat.
  - c. Ambil *K* terdekat ini sesuai dengan jarak *Euclidean* yang di hitung.
  - d. Di antara *K* terdekat ini, hitunglah jumlah titik data dalam setiap kategori.
  - e. Tetapkan titik data baru ke dalam kategori yang jumlah tetangganya maksimum.
  - f. Model yang akan di buat sudah siap.

Kelebihan *K-Nearest Neighbors* adalah algoritmanya yang sederhana dan mudah diimplementasikan serta pelatihan sangat cepat, sederhana dan

mudah dipelajari, tahan terhadap data pelatihan yang memiliki derau dan efektif jika data pelatihan besar. Kekurangan *K-Nearest Neighbors*.

- a. Rentan terhadap variabel yang non-informasi.
- b. Rentan terhadap dimensionalitas yang tinggi.
- c. Rentan terhadap perbedaan rentang variabel.
- d. Nilai komputasi yang tinggi.

2. Rumus Akurasi (*Accuracy*)

Akurasi mengukur persentase citra yang diklasifikasikan dengan benar dari seluruh citra uji yang ada. Rumus akurasi dinyatakan sebagai berikut :

$$Accuracy = \frac{\text{jumla citra benar diklasifikasi}}{\text{jumlah total citra uji}} \dots\dots\dots 1.1$$

3. *Precision*

*Precision* mengukur presentase citra positif yang benar terklasifikasi dari seluruh citra yang diklasifikasi sebagai positif. Rumus *precision* dinyatakan sebagai berikut:

$$Precision = \frac{True\ Positives}{True\ Positives + Fals\ Positives} \dots\dots\dots 1.2$$

4. *Recal* (*Sensitivity* atau *True Positive Rate*)

*Recall* mengukur presentase citra positif yang benar terklasifikasi dari seluruh citra positif yang sebenarnya. Rumus *recall* dinyatakan sebagai berikut:

$$Accuracy = \frac{True\ Positives}{True\ Positives + Fa\ Negatives} \dots\dots\dots 1.3$$

### 5. *F1-Score*

*F1-score* adalah rata-rata harmonik antara *precision* dan *recall*, memberikan keseimbangan antara keduanya. Rumus *F1-Score* dinyatakan sebagai berikut.

$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Precision \times Recall} \dots\dots\dots 1.4$$

### 2.3. *Data mining*

*Data mining* adalah proses ekstraksi pengetahuan atau informasi yang bermanfaat dari sejumlah besar data yang tidak terstruktur atau berantakan. Tujuan utama dari *Data mining* adalah untuk mengidentifikasi pola, hubungan, dan informasi yang tersembunyi dalam data yang dapat digunakan untuk membuat keputusan yang lebih baik dan mendukung pengambilan keputusan dalam berbagai bidang (*Neighbors et al.*, 2023).

Proses *data mining* melibatkan serangkaian langkah dan teknik analisis data yang kompleks, termasuk *preprocessing* data, pemodelan, evaluasi model dan interpretasi hasil. Beberapa teknik umum yang digunakan dalam *Data mining* termasuk *K-Nearest Neighbors* (KNN), *Decision Trees*, *Association Rule Mining*, *Clustering*, dan *Neural Networks* (*Arifuddin et al.*, 2023).

Penggunaan *data mining* meluas di berbagai industri dan bidang, termasuk bisnis, kesehatan, pemasaran, keuangan, ilmu sosial, dan banyak lagi. Beberapa

contoh penerapan data *mining* adalah untuk prediksi perilaku pelanggan, analisis pola pembelian, identifikasi anomali, analisis data kesehatan, pengenalan pola dalam citra dan teks, serta analisis sentimen media sosial. Data *mining* membantu organisasi dan perusahaan untuk menggali wawasan baru dari data yang mereka miliki, membantu dalam pengambilan keputusan yang lebih tepat dan mendukung perencanaan strategis. Selain itu, data *mining* juga merupakan salah satu cabang penting dalam analisis data dan kecerdasan buatan yang terus berkembang dan berkontribusi pada kemajuan teknologi dan pengetahuan (Priatmojo et al., 2023).

#### **2.4 Orange**

*Orange* adalah perangkat lunak *open-source* untuk analisis data, visualisasi, dan pembelajaran mesin. Dikembangkan dengan bahasa pemrograman *Python*, *Orange* menyediakan lingkungan visual yang interaktif dan intuitif, sehingga memungkinkan pengguna untuk melakukan analisis data yang kompleks tanpa harus menulis kode secara manual (Neighbors et al., 2023). Fitur utama *Orange* meliputi (Putra & Fathoni, 2023) :

1. Visualisasi Data: *Orange* menyediakan berbagai jenis visualisasi data, seperti *plot*, *scatterplot*, *heatmap*, dan lainnya, yang membantu pengguna memahami data dengan lebih baik.
2. Data: *Orange* memungkinkan pengguna membersihkan dan mempersiapkan data sebelum dilakukan analisis. Pengguna dapat melakukan manipulasi data, menghapus data yang hilang, dan mengubah format data.

3. Analisis Data: *Orange* menyediakan berbagai algoritma analisis data, termasuk klasifikasi, *clustering*, regresi, dan analisis asosiasi. Pengguna dapat menerapkan algoritma ini pada data mereka dan melihat hasilnya dalam bentuk visual yang interaktif.
4. Pembelajaran Mesin: *Orange* memiliki koleksi algoritma pembelajaran mesin yang lengkap, memungkinkan pengguna untuk melatih model *machine learning* dan melakukan prediksi pada data baru.
5. *Workflows*: *Orange* memungkinkan pengguna untuk membuat *workflows* analisis data yang kompleks dengan menggabungkan berbagai alat dan algoritma. Hal ini mempermudah pengguna dalam mengorganisasi dan mengulangi proses analisis data dengan langkahlangkah yang telah ditentukan.
6. Integrasi *Python*: *Orange* memungkinkan pengguna yang mahir dalam pemrograman *Python* untuk mengintegrasikan kode *Python* dan mengakses fitur-fitur *Orange* secara lebih mendalam.