

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian terkait tentang klasifikasi citra bunga *euphorbia* menggunakan mesin learning sudah dilakukan oleh beberapa peneliti terdahulu. Pada penelitian sebelumnya Diana Theresia Worung yang berjudul "Implementasi *K-NN* Pada Pengklasifikasian Citra Bunga". Hasil penelitiannya menunjukkan bahwa pengujian dengan menggunakan metode *K-NN* tanpa segmentasi memperoleh hasil tingkat akurasi 77%. Sedangkan dengan menggunakan segmentasi memperoleh tingkat akurasi 85% (Worung, 2019).

Pada penelitian lain menurut Nofitha Sari yang berjudul "Implementasi Algoritma *K-NN* Untuk Identifikasi Bunga Anggrek". Hasil penelitiannya menunjukkan bahwa hasil identifikasi dari setiap kelopak bunga anggrek memiliki presentase 86.7% dengan rasio dan training dan testing 45 : 15 (Sari & Wulanningrum, 2021).

Pada penelitian lain menurut Rafie yang berjudul "Klasifikasi Bunga Menggunakan *Naive Bayes* Berdasarkan Fitur Warna Dan Teksture". Hasil penelitiannya menunjukkan bahwa percobaan pertama klasifikasi *Naive Bayes* fitur tingkat pertama akurasinya 66%, percobaan kedua dengan distribusi kernel akurasinya 73%, sedangkan dengan menerapkan 10 fitur dengan distribusi kernel tingkat akuarasinya lebih besar yaitu 77%. Klasifikasi bunga sesuai warna dan bentuk pada nilai AUC sebesar 88% dan menghasilkan akurasi sebesar 71,1% (Hidayatullah, 2021).

Pada penelitian lain menurut Fansyuri Maulana & Hariansyah yang berjudul “Perkenalan objek bunga dengan *ekstraksi fitur* warna dan bentuk menggunakan metode Morfologi dan *Naïve Bayes*.” Hasil penelitiannya menunjukkan bahwa nilai akurasi klasifikasi dengan nilai AUC sebesar 88% dan nilai akurasi 71,1% sehingga pengenalan objek citra sesuai dengan citra sebenarnya (Maulana Fansyuri & Hariansyah, 2020).

Penelitian ini merujuk ke penelitian yang dilakukan oleh Nofitha sari yang berjudul “Implementasi Algoritma *K-NN* Untuk Identifikasi Bunga Anggrek”. Hasil penelitiannya menunjukkan bahwa hasil identifikasi dari setiap kelopak bunga anggrek memiliki presentase 86.7% dengan rasio data *training and testing* 45 : 15.

Perbedaan dari penelitian ini yaitu penelitian ini menggunakan citra bunga *euphorbia* sebagai objek penelitian dan fokusnya terhadap pengelompokan warna bunga sedangkan penelitian Nofitha Sari menggunakan bunga anggrek sebagai objek penelitiannya dan fokusnya terhadap pengelompokan kelopak bunga.

Dan juga merujuk pada penelitian yang dilakukan oleh Rafie yang berjudul “Klasifikasi Bunga Menggunakan *Naive Bayes* Berdasarkan Fitur Warna Dan Teksture”. Hasil penelitiannya menunjukkan bahwa hasil klasifikasi bunga berdasarkan warna dan bentuk dengan nilai AUC sebesar 88% dan menghasilkan akurasi sebesar 71,1%. Perbedaan dari penelitian ini yaitu penelitian ini menggunakan metode *Naive Bayes* untuk mengklasifikasi dan mengukur tingkat akurasi warna bunga *euphorbia* milli sedangkan penelitian Rafie mengklasifikasi fitur warna dan bentuk untuk pengenalan objek bunga.

Tabel 2.1 Perbandingan penelitian terdahulu.

No	Nama (Tahun)	Judul	Metode	Hasil
1	Woung (2020).	Implementasi <i>KNN</i> Pada pengklasifikasian citra bunga.	Metode <i>K-NN</i> .	Pengujian dengan menggunakan metode <i>K-Nearest Neighbor</i> tanpa segmentasi mendapatkan hasil tingkat akurasi 77%. Sedangkan dengan menggunakan segmentasi mendaptkan tingkat akurasi 85%.
2	Sari &Wulann ingrum (2021).	Implementasi algoritma <i>K-nearest neighbor</i> untuk identifikasi bunga anggrek.	Metode Algoritma <i>K-NN</i> .	hasil identifikasi dari setiap kelompok bunga anggrek memiliki prosentase 86.7% menggunakan rasio

				dan training dan testing 45 : 15.
3	Hidayatul lah (2021).	Klasifikasi bunga menggunakan <i>Naïve bayes</i> berdasarkan fitur warna dan texture.	Metode <i>Naïve Bayes</i> .	Percobaan pertamaklasifikasi <i>naïve buyes</i> fitur tingkat pertama akurasinya 66%, percobaan kedua dengan distribusi kenel akurasinya 73%, sedangkan dengan menerapkan 10 fitur dengan distribusi kenel tingkat akuarasinya lebih besar yaitu 77%.

4	Maulana Fansyuri & Hariansyah, (2020).	Pengenalan objek bunga dengan ekstraksi fitur warna dan bentuk menggunakan metode Morfologi dan <i>Naïve Bayes</i> .	Metode Morfologi dan <i>Naïve Bayes</i> .	Nilai akurasi klasifikasi dengan nilai AUC sebesar 88% dan nilai akurasi 71,1% sehingga pengenalan objek citra sesuai dengan citra sebenarnya.
---	--	--	---	--

2.2 Teori Pendukung

2.2.1 Citra Bunga *Euphorbia Milli*

Dalam penelitian ini digunakan 4 jenis pengelompokan warna bunga *euphorbia milli* yang diwakili citra bunga yang berjumlah 2,832 sample yang terdiri dari beberapa sampel warna diantaranya yaitu 708 merah, 708 putih, 708 merah muda dan 708 kuning. Sampel warna bunga tersebut kemudian dilakukan tahapan preprocessing penskalaan ulang pada masing masing citra tersebut sedemikian rupa sehingga matriks representasi dari setiap citra bunga dan sampel berukuran sama.

Setelah itu kemudian selanjutnya citra bunga dan sampel yang berukuran sama diubah menjadi vector-vektor kolom. Kemudian itu akan dilakukan transformasi vector-vektor kolom kedalam ruang berdimensi 2 dan 3 dengan menggunakan metode *K-NN* (Rosyani & Hariansyah, 2020).

2.2.2 Citra Digital

Citra digital adalah representative pada citra yang diperoleh melalui mesin dengan bentuk pendekatan sesuai sampling dan kuantisasi. Citra digital merupakan kumpulan piksel – piksel atau titik – titik yang berwarna yang berbentuk dua dimensi. Sampling meenginformasikan ukuran kotak-kotak yang tersusun pada baris dan kolom. Atau dalam pengertian lain sampling dalam citra meenginformasikan besar kecilnya ukuran pixel (titik) pada citra, dan kuantisasi menyatakan besarnya nilai tingkat kecerahan yang dinyatakan dalam nilai tingkat keabuan sesuai dengan jumlah bit biner yang digunakan oleh mesin, dengan kata lain kuantisasi pada citra menyatakan jumlah warna yang ada pada citra (Rosyani & Hariansyah, 2020).

2.2.3 Klasifikasi *K-NN*

K-NN Classifier didasari pada pembelajaran dengan persamaan sampel yang dibuat dalam penguraian oleh dimensi atribut angka. Masing-masing sampel merepresentasikan sebuah titik pada ruang dimensi. Ketika ada sebuah sampel yang tidak diketahui, *K-NN Classifier* mencari pola ruang dari sampel yang diuji yang sama sampel tersebut dicari yang terdekat dengan sampel yang tidak diketahui.

Terdekat artinya jarak kedekatan data berdasarkan aturan *Euclidean Distance*. Aturan pada *Euclidean Distance* dihitung diantara dua titik, dan setiap atribut telah mendapat bobot nilainya. *K-NN* merupakan metode klasifikasi yang tangguh terhadap training data yang memiliki banyak noise dan metode ini memiliki tingkat akurasi yang cukup tinggi serta efektif pada nilai training datanya besar.

Sesuai dengan prinsip kerja *K-NN* yaitu mencari jarak terdekat antara data yang akan dievaluasi dengan *K* tetangga terdekatnya dalam pelatihan (Tarigan et al.,

2022). Algoritma *K-NN* ditemukan oleh Fix dan Hodges tahun 1951, dan berhasil menjadi algoritma klasifikasi sederhana yang akurat.

K-NN merupakan sebuah cara yang digunakan agar dalam membuat klasifikasi dengan memakai algoritma *supervised*. Manfaat algoritma ini adalah melakukan klasifikasi obyek baru sesuai jarak suatu obyek yang hendak diklasifikasikan bagi data contoh. *Classifier* hanya memakai fungsi jarak dari data baru ke data training.

K-NN merupakan algoritma yang menentukan sendiri nilai jarak yang terdapat pengujian data testing dengan data training yang diambil dari nilai paling kecil dari nilai ketetanggaan terdekat. Dalam penelitian ini data yang dipakai adalah citra bunga *euphorbia milli* yang terdiri dari 4 warna.

Data citra yang digunakan berjumlah 2832 foto citra bunga *euphorbia milli*. Proses klasifikasi citra bunga *euphorbia milli* dengan menerapkan algoritma *K-NN*. Algoritma *K-NN* ditemukan Fix dan Hodges pada tahun 1951, kemudian berhasil sebagai algoritma klasifikasi sederhana yang akurat (Maulana Fansyuri & Hariansyah, 2020).

Jarak dari data yang ada masing-masing dihitung menggunakan persamaan *Euclidean distance* (Jarak *Euclidean*).

$$d(x,y)=\sqrt{\sum_{k=1}^m(x_i - y_i)^2}\dots\dots\dots(1)$$

Dimana:

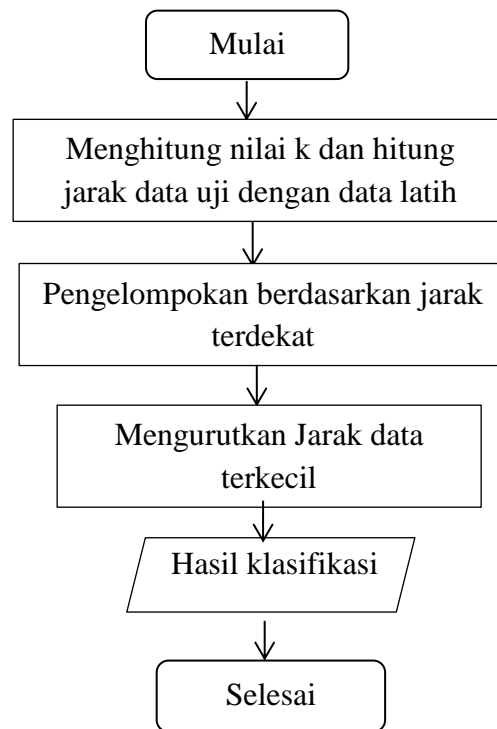
d=jarak data *testing* ke data *training*

m=Jumlah dimensi (*attributes*)

x_i =Data *testing* ke-i, dengan $i=1,2,\dots,m$

y_i =data *training* ke-i, dengan $i=1,2,\dots,m$

Berikut merupakan tahapan dalam Algoritma *K-NN*.



Gambar 2.1 Tahapan algoritma *K-NN*.

Pada tahapan *training*, algoritma ini cuman melakukan penyimpanan vektor-vektor fitur dan klasifikasi data *training* sample. Pada tahap klasifikasi, fitur-fitur yang sama dihitung untuk testing data (yang klasifikasinya tidak diketahui).

Jarak pada vektor baru terhadap semua vektor *training sample* dihitung melalui sejumlah *k*. hasil terdekat diambil titik terbaru klasifikasinya diprediksikan termasuk dalam klasifikasi yang banyak dari titik-titik tersebut.

2.2.4 Klasifikasi Naive Bayes

Naive Bayes adalah metode yang menggunakan cabang matematika yang dikenal dengan teori probabilitas dalam mencari peluang paling besar dari kemungkinan klasifikasi dengan Teknik melihat frekuensi masing-masing klasifikasi dalam data training.

Naïve bayes adalah metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik pada data mining. Nama lain dari *naïve bayes* yaitu *Idiot's Bayes*, *Simple Bayes* dan *Idependence Bayes*.

Klasifikasi *Naïve Bayes statistic* merupakan pengelompokan yang bias dipakai dalam memprediksi probabilitas keanggotaan suatu *class*. Klasifikasi Bayesian ditentukan pada teorema *Bayes*, diperoleh dari nama seorang ahli matematika yang juga menteri *Prebyterian Inggris*, Klasifikasi Bayesian memiliki kemampuan klasifikasi terhubung dengan *decision tree* dan *neural network* Hayuninghtyas (2019).

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \dots\dots\dots(2)$$

Keterangan :

y = data dengan kelas yang belum diketahui

x = hipotesis data y merupakan satu kelas spesifik

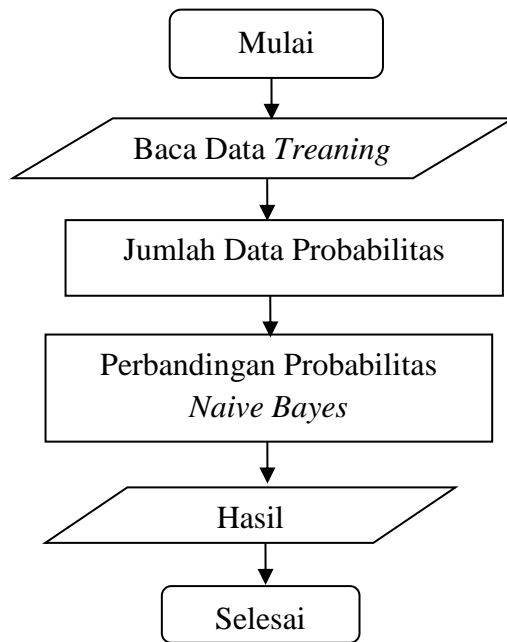
P(x|y) = probabilitas hipotesis x berdasarkan kondisi

P(x) = probabilitas hipotesis x (*prior probability*)

P(y|x) = probabilitas y berdasarkan kondisi pada hipotesis x

P(y) = probabilitas dari y

Berikut tahap dalam algoritma *Naïve Bayes*.



Gambar 2.2 Tahapan Algoritma *Naïve Bayes*.

2.2.5 Confusion Matrix

Confusion Matrix adalah salah satu *tools* yang digunakan untuk mengevaluasi dalam memprediksi objek yang benar dan salah (Ardiansyah & Wilma, 2018). Suatu matrix prediksi sebagai pembanding antara asli dari inputan atau berupa informasi nilai aktual dan prediksi dalam klasifikasi. Tingkatan akurasi yaitu tingkat akurasi jaringan yang telah dibuat dalam mendeteksi masukan citra yang diberikan sehingga menghasilkan *output* yang benar.

Confusion matrix yang dihitung dengan rumus:

$$a) \text{ Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \dots\dots\dots(1)$$

$$b) \text{ Precision} = \frac{TP}{TP+FP} \dots\dots\dots(2)$$

$$c) \text{ Recall} = \frac{TP}{TP+FN} \dots\dots\dots(3)$$

$$d) \text{ FI} = \frac{\text{Recall}+\text{Precision}}{\text{Recall}+\text{Precision}} \times 2 \dots\dots\dots(4)$$

Keterangan :

TN = Merupakan data positif yang diprediksi benar

TN = Merupakan data negative yang diprediksi benar

FP = Memprediksi positif dan itu salah

FN = Memprediksi negative dan itu salah

1. *Test and Score*

Evaluasi hasil model dengan mendapatkan data numerik.

a. *Area under the curve (AUC)*

Prediksi yang terdapat dalam membandingkan model *AUC* memudahkan dalam membandingkan model satu dengan yang lainnya.

b. *Classification Accuracy (CA)*

Menghitung tingkat akurasi klasifikasi.

c. F1

Rata-rata terimbang dari presisi dan daya ingat dengan rumus yang di hitung pada persamaan 4.

d. *Precision*

Kemampuan *classifier* untuk tidak memberi label positif kepada sampel negative begitupun sebaliknya. Precision memiliki rumus yang di hitung pada persamaan 2.

e. *Recall*

Kemampuan *classifier* untuk menemukan dan menggolongkan semua sampel yang bernilai positif.