

# Pengenalan Objek Bunga dengan Ekstraksi Fitur Warna dan Bentuk Menggunakan Metode Morfologi dan Naïve Bayes

Maulana Fansyuri<sup>1</sup>, Oke Hariansyah<sup>2</sup>

Universitas Pamulang

e-mail: <sup>1</sup>dosen02359@unpam.ac.id, <sup>2</sup>dosen00840@unpam.ac.id

Diajukan: 6 September 2020; Direvisi: 25 November 2020; Diterima: 30 November 2020

## Abstrak

*Pengenalan objek pada citra digital mendapat banyak perhatian dalam beberapa tahun terakhir ini. Hal ini terlihat dari penerapan identifikasi citra di beberapa bidang kerja seperti pengenalan gambar untuk perencanaan lahan pertanian, identifikasi penyakit tanaman dan pencitraan medis. Teknik pengolahan gambar juga banyak digunakan oleh ahli botani dalam mengidentifikasi spesies tanaman yang tidak diketahui. Metode yang di terapkan di dalam penelitian ini untuk masalah proses pengenalan citra bunga adalah Ekstraksi model warna dan ekstraksi fitur bentuk untuk menaikkan tingkat akurasi dalam pengenalan citra bunga. Hasil ekstraksi warna dan bentuk bunga tersebut diklasifikasikan menggunakan metode Naïve Bayes dengan tujuan agar bisa mendapatkan tingkat akurasi terbaik dalam proses klasifikasi citra bunga. Penelitian ini menggunakan 285 data citra bunga yang merupakan data set dari 19 spesies bunga yang berbeda, dataset dalam penelitian ini dibagi menjadi dua bagian, yaitu 228 data tes dan 57 data training. Hasil dari ekstraksi fitur dan warna menggunakan metode Morfologi dan Naïve Bayes ini dilakukan pengujian akurasi menggunakan metode Confusion Matrix, menghasilkan tingkat akurasi sebesar 71,1% dengan kategori klasifikasi yang baik.*

**Kata kunci:** *Pengenalan, Klasifikasi, Naïve Bayes, Confusion Matrix.*

## Abstract

*The introduction of objects in digital images has received a lot of attention in recent years. This can be seen from the application of image identification in several fields of work such as image recognition for agricultural land planning, identification of plant diseases and medical imaging. Image processing techniques are also widely used by botanists in identifying unknown plant species. The method applied in this research for the problem of the flower image recognition process is the extraction of the color model and the extraction of shape features to increase the level of accuracy in the recognition of flower images. The results of the color and flower shape extraction are classified using the Naïve Bayes method in order to get the best level of accuracy in the flower image classification process. This study uses 285 flower image data which is a data set of 19 different flower species. The dataset in this study is divided into two parts, namely 228 test data and 57 training data. The results of the feature and color extraction using the Morphology and Naïve Bayes methods were accuracy tested using the Confusion Matrix method, resulting in an accuracy rate of 71.1% with a good classification category.*

**Keywords:** *Recognition, Classification, Naïve Bayes, Confusion Matrix.*

## 1. Pendahuluan

Pengenalan objek pada citra digital menjadi bidang ilmu yang banyak diteliti akhir-akhir ini. Hal ini terlihat dari penerapan di beberapa bisa seperti pengenalan gambar untuk perencanaan lahan pertanian [1][2], pengenalan citra wajah [3], dan pencitraan medis[4]. Teknik pengolahan gambar juga banyak digunakan oleh ahli botani dalam mengidentifikasi spesies tanaman yang tidak diketahui. Meskipun daun sangat penting di dalam proses pengenalan tanaman, bagian lain seperti bunga dapat mendukung proses pengenalan tersebut [5].

Klasifikasi bunga dalam proses pengenalan citra bunga adalah proses yang cukup menantang seperti yang ditunjukkan oleh penelitian Rosyani, dkk [5]. Di dalam penelitiannya mengombinasikan fitur warna RGB, HSV, LAB, dan YCBCR dengan kombinasi segmentasi K-Means dan klasifikasi SVM untuk mendapatkan tingkat akurasi yang baik. Sistem pengenalan bunga juga diteliti oleh Apriyanti, dkk [6] yang meneliti tentang Bunga Angrek.

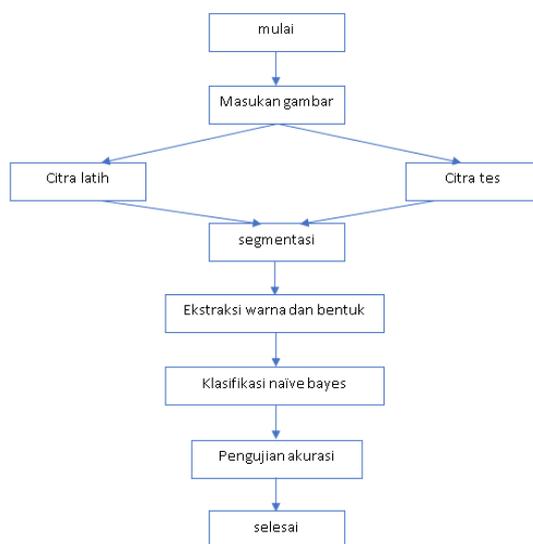
Di dalam proses klasifikasi hal yang perlu dilakukan sebelumnya adalah mengambil fitur-fitur yang dibutuhkan seperti fitur bentuk dan warna pada bunga. Pengambilan fitur ini dilakukan untuk mengambil daerah penting daerah penting ini adalah memisahkan objek dengan *background* pada citra, dibutuhkan proses pengambilan fitur atau ekstraksi fitur. Untuk mengetahui data dari suatu citra, diperlukan metode ekstraksi ciri. Ekstraksi ciri merupakan metode dalam pemrosesan citra dengan mengambil beberapa bagian pada citra yang bisa menunjukkan ciri khas dari citra tersebut, data yang dapat diambil misalnya pola, warna, bentuk, diameter, dan masih banyak lagi. Ekstraksi citra warna yaitu data citra digital yang mempunyai parameter kombinasi data warna *red*, *green*, dan *blue*, sedangkan parameter dihasilkan dari Ekstraksi citra bentuk yaitu *eccentricity* dan *metric*.

Penggunaan metode Naïve Bayes sudah banyak diterapkan untuk melakukan klasifikasi data hasil ekstraksi data citra, pada penelitian ini penggunaan metode Naïve Bayes bertujuan untuk bisa mendapatkan tingkat akurasi terbaik dalam proses klasifikasi citra bunga, sehingga memperkuat kesesuaian data hasil pengolahan citra dengan data sebenarnya. Penerapan metode Naïve Bayes sudah pernah dilakukan untuk mengklasifikasikan citra buah belimbing, dengan metode ini tingkat akurasi yang didapatkan sebesar 80% [7]. Dari hasil tersebut, peneliti tertarik menggunakan metode Naïve Bayes untuk menerapkan klasifikasi citra berdasarkan warna dan bentuk pada bunga.

Berdasarkan uraian di atas, maka penelitian ini menganalisis citra bunga di mana kombinasi yang digunakan adalah dari data warna yang terdiri dari fitur HSV dan YCBCR, dan data bentuk yang terdiri dari fitur *eccentricity* dan *metric*. Hasil ekstraksi warna dan bentuk tersebut dianalisis menggunakan algoritma Naïve Bayes sehingga keakuratan data ekstraksi dengan data sebenarnya lebih optimal. Penelitian ini dapat digunakan oleh ahli botani atau petani untuk mendapatkan teori dan produksi bunga yang lebih baik.

## 2. Metode Penelitian

Metode dalam penelitian ini dilakukan dalam dua fase, yaitu fase pelatihan dan fase pengujian. Dalam fase pelatihan, tahapan yang dilakukan bertujuan untuk mendapatkan model berdasarkan *subset* citra yang disebut citra latih [8]. Ada beberapa tahapan dalam proses Analisis citra bunga berdasarkan warna dan bentuk, tahapan tersebut termuat dalam diagram alur seperti di bawah ini:



Gambar 1. Metode penelitian.

Berikut penjelasan dari tahapan metode penelitian yang dilakukan:

- Langkah pertama adalah menyiapkan data set gambar yang akan dianalisis. *Dataset* gambar yang digunakan adalah 19 gambar bunga dari *dataset* ImageClef2018 dan melakukan pengambilan gambar sebanyak 285 kali. Gambar bunga yang diambil tersebut memiliki *background* yang kompleks sehingga untuk proses selanjutnya harus memisahkan *background* dan objek (*foreground*) dari gambar tersebut.
- Dari 285 *dataset* ini kemudian dibagi menjadi 2 bagian yaitu 80% akan digunakan sebagai data tes yaitu sebanyak 228 data citra dan 20% akan digunakan sebagai data latih yaitu sebanyak 57 data citra.

3. Dari setiap gambar latih dan tes tersebut akan dilakukan segmentasi untuk memisahkan objek (*foreground*) dan *background* menggunakan metode K-Means *Clustering*. Tahapan proses segmentasi menggunakan metode ini yaitu [9].
  - a. Membaca citra asli
  - b. Mengonversi RGB dari citra tersebut menjadi *grayscale*
  - c. Melakukan segmentasi citra menggunakan algoritma K-Means *Clustering* yang bertujuan untuk membagi dua region *cluster* yaitu kluster 1 untuk region berwarna hitam yang menunjukkan *background* data gambar tersebut dan kluster 2 untuk region berwarna kuning yang menunjukkan objek (*foreground*) dari gambar tersebut.
  - d. Memilih region kluster yang memiliki luas paling kecil untuk menunjukkan bahwa region tersebut adalah objek (*foreground*) dari data yang akan diolah.
  - e. Melakukan proses *filtering* dengan menggunakan median filter pada citra tersebut.
  - f. Melakukan proses operasi morfologi dengan tujuan untuk menghilangkan *noise* pada hasil *filtering* sehingga mendapatkan hasil segmentasi yang lebih akurat.
  - g. Mendapatkan data citra biner hasil operasi morfologi tersebut
4. Setelah data citra biner didapatkan dari hasil segmentasi, proses selanjutnya adalah melakukan ekstraksi warna dan bentuk dari data tersebut, di mana data yang diambil dari fitur warna adalah data rata-rata dari HSV dan data rata-rata dari YCBCR, serta fitur bentuk yang diambil adalah dari *eccentricity* dan *metric*.
5. Langkah selanjutnya adalah penerapan metode Naïve Bayes dari data fitur warna dan bentuk yang telah didapat. Tahapan proses metode Naïve Bayes ini yaitu [7]:
  - a. Menghitung nilai probabilitas (*prior*) dari setiap variabel dari data tes yang didapat dari hasil ekstraksi citra.
  - b. Karena data hasil Ekstraksi citra merupakan data numerik, maka langkah selanjutnya adalah mencari nilai rata-rata (*mean*) dari setiap variable dengan persamaan:

$$\mu = \sum_{i=1}^n xi \tag{1}$$

- c. Kemudian menghitung standar deviasi dari setiap variabel tersebut dengan persamaan.

$$\sigma = \sqrt{\sum_{i=1}^n \frac{(xi-\mu)^2}{2}} \tag{2}$$

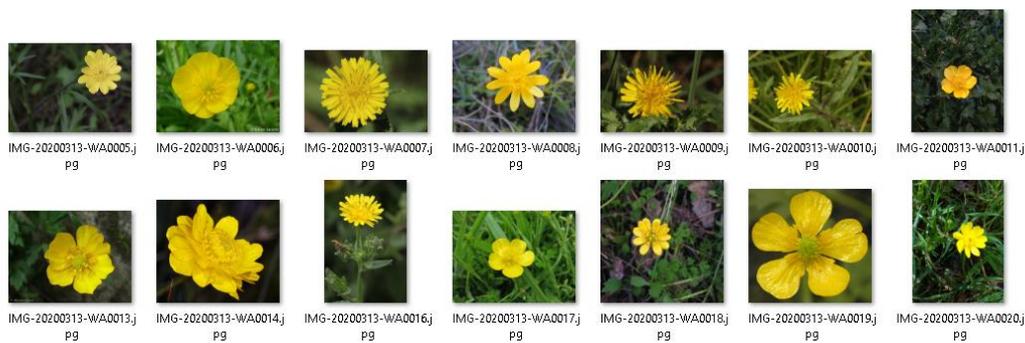
- d. Kemudian menghitung Densitas probabilitas dari setiap jenis spesies citra bunga dengan mengambil sampel dari data citra latih untuk menentukan keakuratan data berdasarkan data tes yang sudah diolah dengan persamaan:

$$p(x = v | c) = \frac{1}{\sqrt{2\mu\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2} \tag{3}$$

- e. Probabilitas yang dihasilkan dari perhitungan densitas probabilitas dan memiliki nilai paling besar akan menunjukkan spesies dari citra tersebut.
- f. Setelah dilakukan pengolahan data menggunakan algoritma Naïve Bayes, akan didapatkan nilai akurasi data hasil ekstraksi dengan data citra bunga sebenarnya.

### 2.1. Data Citra Collection

*Dataset* yang digunakan terdiri dari 19 jenis spesies citra bunga dan masing-masing spesies memiliki 15 citra bunga sehingga total data citra yang digunakan sebagai *dataset* berjumlah 285 data. Dari seluruh *dataset* ini kemudian dibagi menjadi 2 bagian yaitu 80% akan digunakan sebagai data tes yaitu sebanyak 228 data citra dan 20% akan digunakan sebagai data latih yaitu sebanyak 57 data citra.



Gambar 2. Contoh dataset.

### 2.2. Image Data Pre-Processing

Pada tahapan ini merupakan tahap segmentasi citra menggunakan metode K-Means *Clustering*, pada segmentasi ini dilakukan menggunakan aplikasi MATLAB untuk dapat melakukan ekstraksi citra bunga menjadi data citra biner yang dapat menghasilkan ekstraksi warna dan bentuk. Tahap awal dalam segmentasi ini yaitu membagi citra bunga menjadi dua region *cluster* yaitu *cluster* hitam yang menunjukkan *background* dan *cluster* kuning yang menunjukkan objek atau *foreground*.

Selanjutnya memilih region kluster yang paling luas pada gambar yang menunjukkan bahwa region tersebut adalah objek atau *foreground*, sehingga hasil seleksi ini dapat diolah di langkah selanjutnya. Setelah melakukan seleksi objek, langkah selanjutnya adalah melakukan proses *filtering* dengan menggunakan median filter pada citra. Selanjutnya adalah melakukan proses morfologi yang bertujuan untuk menghilangkan *noise* pada gambar yang sudah di filter sehingga dapat memberikan hasil segmentasi yang lebih akurat. Tahap akhir dari segmentasi ini adalah mendapatkan data citra biner dari hasil operasi morfologi citra .

### 2.3. Citra Training

Citra *training* merupakan data hasil segmentasi menggunakan metode K-Means *Clustering* dan ekstraksi citra. Data ini berisi fitur warna yang terdiri dari nilai *Mean H*, *Mean S*, *Mean V*, *Mean Y*, *Mean CB*, *Mean CR*, dan fitur bentuk dari citra bunga yang terdiri dari *eccentricity* dan *metric*.

### 2.4. Metode Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi yang sederhana di mana setiap atribut bersifat independen dan memungkinkan berkontribusi terhadap keputusan akhir. Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Persamaan dari Teorema Bayes adalah:

$$p(H|X) = \frac{p(H|X) \cdot P(X)}{P(X)} \tag{4}$$

Di mana:

- X* : Data dengan *class* yang belum diketahui
- H* : Hipotesis data *X* merupakan suatu *class* spesifik
- P(H|X)* : Probabilitas hipotesis *H* berdasar kondisi *X* (*posterior probability*)
- P(H)* : Probabilitas hipotesis *H* (*prior probability*)
- P(X|H)* : Probabilitas *X* berdasarkan kondisi pada hipotesis *H*
- P(X)* : Probabilitas *X*

Untuk menjelaskan teorema Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema Bayes di atas disesuaikan sebagai berikut:

$$P(C|F1 \dots Fn) = \frac{P(C) \cdot P(F1 \dots Fn|C)}{P(F1 \dots Fn)} \tag{5}$$

Di mana variabel  $C$  mempresentasikan kelas, sementara variabel  $(F_1 \dots F_n)$  mempresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas  $C$  (Posterior) adalah peluang munculnya kelas  $C$  (sebelum masuknya sampel tersebut, sering kali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas  $C$  (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*).

Adapun alur dari metode Naïve Bayes adalah sebagai berikut:

1. *Input* data
2. *Training* baca data *training*.
3. Hitung jumlah dan probabilitas, namun apabila data numerik maka:
  - a. Cari nilai *mean* dan standar deviasi dari masing-masing parameter yang merupakan data numerik.
  - b. Cari nilai probabilitas dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
4. Mendapatkan nilai dalam tabel *mean*, standar deviasi, dan probabilitas.

### 3. Hasil dan Pembahasan

Ada beberapa tahapan dalam proses Analisis citra bunga berdasarkan warna dan bentuk yang dilakukan dalam penelitian ini.

#### 3.1. Persiapan Gambar

Data set yang digunakan terdiri dari 19 jenis spesies citra bunga dan masing-masing spesies memiliki 15 citra bunga sehingga total data citra yang digunakan sebagai *dataset* berjumlah 285 gambar.

#### 3.2. Pembagian Dataset

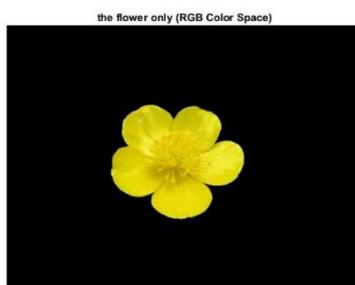
Dari seluruh *dataset* ini kemudian dibagi menjadi 2 bagian yaitu 80% akan digunakan sebagai data latih yaitu sebanyak 228 data citra dan 20% akan digunakan sebagai data tes yaitu sebanyak 57 data citra.

#### 3.3. Proses Segmentasi

Tahap awal segmentasi ini dilakukan pada fase pelatihan yaitu tahapan yang dilakukan bertujuan untuk mendapatkan model berdasarkan citra latih. Tahap segmentasi citra menggunakan metode *K-Means Clustering*, pada segmentasi ini dilakukan menggunakan aplikasi MATLAB untuk dapat melakukan ekstraksi citra bunga menjadi data citra biner yang dapat menghasilkan ekstraksi warna dan bentuk. Proses awal segmentasi ini diterapkan pada data tes, yaitu citra bunga sebanyak 57 data.

Selanjutnya *preprocessing*, yaitu melakukan operator-operator seperti konversi, *holes filling*, open area. pemilihan bunga jika ada bunga yang lebih dari satu, untuk *case* gambar bunga lebih dari satu diperlukan metode khusus seperti *merging* untuk memisahkan bunga agar hanya terdapat 1 bunga saja di dalam *frame*.

Setelah tampil 1 bunga dalam *frame*, selanjutnya membagi citra bunga menjadi dua region *cluster* yaitu *cluster* hitam yang menunjukkan *background* dan *cluster* kuning yang menunjukkan objek atau *foreground*.



Gambar 3. Contoh gambar hasil segmentasi citra.

#### 3.4. Ekstraksi Warna dan Bentuk

Selanjutnya adalah melakukan proses ekstraksi warna dan bentuk menggunakan metode morfologi yang bertujuan untuk menghilangkan *noise* pada gambar yang sudah di filter sehingga dapat memberikan hasil segmentasi yang lebih akurat. Tahap akhir dari segmentasi ini adalah mendapatkan data citra biner dari hasil operasi morfologi citra.

Hasil dari ekstraksi citra, selanjutnya dibuat data citra *training* yang merupakan data hasil segmentasi menggunakan metode K-Means *Clustering* dan ekstraksi citra yang telah disederhanakan agar proses Analisis menggunakan metode Naïve Bayes lebih terfokus. Data ini berisi fitur warna yang terdiri dari nilai *Mean H, Mean S, Mean V, Mean Y, Mean CB, Mean CR*, dan fitur bentuk dari citra bunga yang terdiri dari *eccentricity* dan *metric*.

Setelah selesai proses ekstraksi warna dan bentuk menggunakan data tes, selanjutnya adalah fase pengujian, yaitu tahapan yang dilakukan bertujuan untuk mendapatkan model berdasarkan citra tes. Data yang diproses berjumlah 228 data citra. Proses yang dilakukan sama seperti yang diterapkan pada data tes, yaitu proses segmentasi hingga ekstraksi warna dan bentuk.

### 3.5. Klasifikasi Naïve Bayes

Variabel yang digunakan dalam perhitungan Naïve Bayes menggunakan fitur warna yang terdiri dari *Mean H, Mean S, Mean V, Mean Y, Mean CB, Mean CR*, dan fitur bentuk yang terdiri dari *eccentricity* dan *metric* yang merupakan hasil ekstraksi citra bunga. Data ekstraksi citra hasil data fase pelatihan dan fase pengujian yang berjumlah 228 untuk data tes dan 57 untuk data latih.

Tabel 1. Contoh data latih.

Spesies	HSV MEAN			Pola - eccentricity		Y CB CR MEAN		
	H_mean	S_mean	V_mean	eccentricity	metric	Y_mean	Cb_mean	Cr_mean
Brodiaea elegans Hoover	0,708	0,469	0,754	0,253	0,294	122,820	160,732	137,507
Brodiaea elegans Hoover	0,724	0,447	0,799	0,100	0,185	134,956	158,640	143,362
Brodiaea elegans Hoover	0,783	0,469	0,693	0,363	0,175	128,662	148,404	152,961
Calochortus kennedyi Porter	0,046	0,871	0,800	0,323	0,788	107,410	88,677	189,190
Calochortus kennedyi Porter	0,166	0,840	0,391	0,414	0,873	50,415	112,742	165,174
Calochortus kennedyi Porter	0,084	0,825	0,846	0,628	0,702	139,637	73,152	172,926
Calochortus luteus Douglas ex	0,155	0,936	0,311	0,360	0,802	73,810	96,541	135,498
Calochortus luteus Douglas ex	0,154	0,957	0,311	0,266	0,834	73,589	96,227	135,529
Calochortus luteus Douglas ex	0,155	0,946	0,359	0,480	0,622	82,914	91,622	136,484
Calochortus splendens Douglas ex	0,829	0,335	0,808	0,176	0,895	157,827	147,196	151,133
Calochortus splendens Douglas ex	0,821	0,348	0,700	0,262	0,788	137,244	145,176	147,256
Calochortus splendens Douglas ex	0,811	0,534	0,659	0,450	0,680	112,617	154,659	153,910
Dichelostemma congestum (Sm.) Kunth	0,732	0,518	0,818	0,581	0,138	127,127	167,042	139,590
Dichelostemma congestum (Sm.) Kunth	0,740	0,294	0,648	0,851	0,108	129,995	143,724	134,286
Dichelostemma congestum (Sm.) Kunth	0,742	0,345	0,549	0,437	0,196	108,422	143,494	134,512

Selanjutnya menghitung *Mean* dari setiap variabel menggunakan persamaan:

$$\mu = \sum_{i=1}^n xi \tag{6}$$

Menghitung nilai *mean* dari data latih yaitu menghitung rata-rata dari setiap spesies data bunga, misal untuk data spesies *Brodiaea elegans Hoover* memiliki nilai *Hue* yaitu: 0,708, 0,724 dan 0,783. Maka rata-rata dari nilai *Hue* tersebut adalah 0,738. Hasil perhitungan *Mean* seluruh data latih terlihat pada tabel di bawah ini:

Tabel 2. *Mean*

Mean	H	S	V	Eccentricity	Metric	Y	CB	CR
Brodiaea elegans Hoover	0,738	0,462	0,749	0,239	0,218	128,813	155,926	144,610
Calochortus kennedyi Porter	0,099	0,845	0,679	0,455	0,788	99,154	91,524	175,763
Calochortus luteus Douglas ex	0,155	0,947	0,327	0,369	0,753	76,771	94,797	135,837
Calochortus splendens Douglas ex	0,820	0,406	0,722	0,296	0,788	135,896	149,010	150,767
Dichelostemma congestum (Sm.) Kunth	0,738	0,386	0,672	0,623	0,148	121,848	151,420	136,129
Encelia californica Nutt.	0,151	0,807	0,731	0,731	0,328	154,299	66,895	143,383
Hemerocallis fulva (L.) L.	0,088	0,776	0,899	0,375	0,290	151,678	75,878	172,690
Hypoxis hirsuta (L.) Coville	0,152	0,905	0,898	0,420	0,252	182,896	42,368	149,609
Lilium columbianum Leichtlin	0,099	0,753	0,823	0,769	0,503	149,599	76,515	162,032
Lilium pardalinum Kellogg	0,215	0,829	0,530	0,739	0,330	80,201	105,392	165,351
Lilium parvum Kellogg	0,066	0,794	0,757	0,329	0,308	118,234	89,940	174,407
Narcissus pseudonarcissus L.	0,147	0,867	0,770	0,662	0,303	157,693	59,900	147,545
Pyrrhopappus carolinianus (Walter) DC.	0,165	0,743	0,697	0,661	0,194	152,002	74,973	137,354
Tetranneuris scaposa (DC.) Greene	0,136	0,924	0,829	0,553	0,348	160,820	52,653	154,816

Tragopogon dubius Scop.	0,159	0,791	0,818	0,419	0,099	173,782	59,517	142,450
Tragopogon porrifolius L.	0,787	0,421	0,667	0,639	0,099	124,866	138,895	149,959
Trillium erectum L.	0,860	0,701	0,624	0,572	0,272	93,929	127,020	170,913
Tussilago farfara L.	0,153	0,799	0,875	0,517	0,232	182,786	54,560	146,017
Wyethia angustifolia (DC.) Nutt.	0,140	0,940	0,790	0,470	0,145	153,508	54,419	153,074

Setelah nilai *Mean* dari setiap variabel didapat, selanjutnya menghitung standar deviasi dari setiap variable dengan menggunakan persamaan:

$$\sigma = \sqrt{\sum_{i=1}^n \frac{(x_i - \mu)^2}{2}} \tag{7}$$

Hasil perhitungan standar deviasi menggunakan data latih terlihat pada tabel di bawah ini:

Tabel 3. Standar deviasi.

Standar Deviasi	H	S	V	eccentricity	metric	Y	CB	CR
Brodiaea elegans Hoover	0,040	0,013	0,053	0,132	0,066	6,070	6,598	7,802
Calochortus kennedyi Porter	0,061	0,024	0,251	0,157	0,085	45,180	19,948	12,257
Calochortus luteus Douglas ex	0,000	0,010	0,028	0,107	0,114	5,321	2,754	0,560
Calochortus splendens Douglas ex	0,009	0,112	0,077	0,141	0,107	22,635	4,995	3,342
Dichelostemma congestum (Sm.) Kunth	0,005	0,118	0,136	0,210	0,045	11,715	13,529	2,999
Encelia californica Nutt.	0,014	0,105	0,046	0,301	0,168	13,885	6,632	4,096
Hemerocallis fulva (L.) L.	0,037	0,012	0,034	0,072	0,135	14,692	10,710	16,163
Hypoxis hirsuta (L.) Coville	0,007	0,051	0,070	0,168	0,081	12,305	7,007	3,064
Lilium columbianum Leichtlin	0,030	0,091	0,061	0,048	0,035	18,503	18,093	4,759
Lilium pardalinum Kellogg	0,117	0,024	0,094	0,148	0,279	17,759	9,059	3,690
Lilium parvum Kellogg	0,019	0,065	0,003	0,149	0,020	8,430	8,627	6,176
Narcissus pseudonarcissus L.	0,001	0,058	0,037	0,006	0,050	7,889	0,804	0,257
Pyrrhopappus carolinianus (Walter) DC.	0,016	0,143	0,113	0,197	0,106	26,256	3,513	5,422
Tetranuris scaposa (DC.) Greene	0,011	0,035	0,089	0,155	0,064	22,701	8,463	3,506
Tragopogon dubius Scop.	0,010	0,028	0,055	0,019	0,022	8,580	5,107	4,645
Tragopogon porrifolius L.	0,022	0,065	0,163	0,176	0,010	23,282	12,819	11,413
Trillium erectum L.	0,078	0,133	0,187	0,201	0,111	30,778	10,417	8,446
Tussilago farfara L.	0,003	0,028	0,065	0,139	0,209	12,878	8,251	1,863
Wyethia angustifolia (DC.) Nutt.	0,022	0,002	0,083	0,159	0,037	9,588	5,243	10,891

Selanjutnya menghitung densitas probabilitas dari setiap variabel menggunakan persamaan:

$$p(x = v | c) = \frac{1}{\sqrt{2\mu\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2} \tag{8}$$

Hasil perhitungan densitas probabilitas menggunakan data latih dari setiap variabel terlihat pada tabel di bawah ini:

Tabel 4. Densitas probabilitas.

Data Sample	H_mean	S_mean	V_mean	eccentricity	metric	Y_mean	Cb_mean	Cr_mean
Narcissus pseudonarcissus L.	0,1478	0,9010	0,7487	<b>0,6661</b>	<b>0,2746</b>	<b>153,1386</b>	<b>59,4358</b>	<b>147,3963</b>
<b>Densitas Probabilitas</b>								
Brodiaea elegans Hoover	2,6E-48	1,2E-262	1,7E+00	6,0E-03	1,1E+00	5,3E-05	5,6E-48	1,3E-00
Calochortus kennedyi Porter	1,2E+00	1,6E-01	7,7E-01	4,1E-01	1,9E-08	2,9E-02	2,5E-02	7,8E-03
Calochortus luteus Douglas ex	8,8E-113	2,6E-04	1,3E-50	2,7E-02	1,9E-04	3,2E-46	3,8E-37	2,3E-93
Calochortus splendens Douglas ex	0,0E+00	6,3E-05	1,4E+00	3,3E-02	1,3E-05	6,3E-02	2,6E-71	1,3E-01

Dichelostemma congestum (Sm.) Kunth	0,0E+00	7,9E-05	9,2E-01	8,5E-01	3,4E-02	3,3E-03	1,0E-11	2,0E-04	0,0E+00	Dichelostemma congestum (Sm.) Kunth
Encelia californica Nutt.	3,3E+00	8,3E-01	1,7E+00	7,1E-01	9,2E-01	1,1E-01	8,2E-02	1,2E-01	3,3E-03	Encelia californica Nutt.
Hemerocallis fulva (L.) L.	5,7E-01	3,1E-22	1,3E-04	3,7E-04	1,1E+00	1,0E-01	3,8E-02	2,9E-02	1,0E-33	Hemerocallis fulva (L.) L.
Hypoxis hirsuta (L.) Coville	4,1E+00	1,8E+00	1,6E-01	3,3E-01	1,3E+00	6,1E-03	7,8E-03	1,8E-01	4,2E-06	Hypoxis hirsuta (L.) Coville
Lilium columbianum Leichtlin	5,9E-01	3,5E-01	7,6E-01	1,8E-01	1,5E-09	9,1E-02	6,0E-02	1,6E-03	3,8E-16	Lilium columbianum Leichtlin
Lilium pardalinum Kellogg	9,9E-01	2,6E-02	8,4E-02	9,2E-01	7,4E-01	2,1E-05	3,4E-07	1,5E-06	1,6E-20	Lilium pardalinum Kellogg
Lilium parvum Kellogg	3,2E-04	4,1E-01	9,6E-02	8,1E-02	6,8E-01	2,6E-05	2,6E-04	1,1E-05	5,4E-20	Lilium parvum Kellogg
Narcissus pseudonarcissus L.	1,1E+01	1,4E+00	1,8E+00	4,2E+00	1,5E+00	1,2E-01	3,8E-01	6,7E-01	5,4E+00	<b>Narcissus pseudonarcissus L.</b>
Pyrrhopappus carolinianus (Walter) DC.	1,8E+00	5,8E-01	1,1E+00	9,0E-01	9,2E-01	7,8E-02	1,2E-05	3,1E-02	2,6E-08	Pyrrhopappus carolinianus (Walter) DC.
Tetraneuris scaposa (DC.) Greene	2,1E+00	1,7E+00	8,9E-01	7,8E-01	8,1E-01	7,9E-02	9,9E-02	2,3E-02	3,7E-04	Tetraneuris scaposa (DC.) Greene
Tragopogon dubius Scop.	2,1E+00	1,1E-03	7,8E-01	6,7E-36	1,3E-13	7,5E-03	1,8E-01	1,1E-01	2,1E-55	Tragopogon dubius Scop.
Tragopogon porrifolius L.	1,3E-178	2,3E-12	8,7E-01	9,4E-01	3,4E-70	4,0E-02	5,1E-10	1,2E-01	1,9E-271	Tragopogon porrifolius L.
Trillium erectum L.	1,0E-18	3,5E-01	7,4E-01	8,0E-01	1,2E+00	1,1E-02	8,9E-11	2,8E-03	7,4E-34	Trillium erectum L.
Tussilago farfara L.	2,0E+00	2,6E-03	2,3E-01	6,0E-01	8,6E-01	7,9E-03	1,2E-01	2,2E-01	1,3E-07	Tussilago farfara L.
Wyethia angustifolia (DC.) Nutt.	2,5E+00	1,8E-64	1,2E+00	4,7E-01	4,6E-03	1,3E-01	1,1E-01	1,1E-01	1,7E-69	Wyethia angustifolia (DC.) Nutt.
Probabilitas Terbesar									5,3508E+00	<b>Narcissus pseudonarcissus L.</b>

Probabilitas yang dihasilkan dari perhitungan densitas probabilitas dan memiliki nilai paling besar akan menunjukkan spesies dari citra tersebut, dari nilai probabilitas dari tabel di atas, nilai probabilitas terbesar ditunjukkan oleh densitas probabilitas dari spesies “Narcissus pseudonarcissus L” dan hasil ini akurat dengan data *sample* yang digunakan yaitu data tes dari spesies “Narcissus pseudonarcissus L”.

### 3.6. Pengujian Akurasi

Keakuratan hasil klasifikasi dapat diukur dengan menggunakan Confusion Matrix. Confusion Matrix adalah media yang berguna untuk menganalisis seberapa baik *classifier* dapat mengenali *tupel* dari kelas yang berbeda [10]. Misalkan terdapat dua kelas, maka akan diistilahkan sebagai *tupel* positif dan *tupel* negatif. *True positive* mengacu pada *tupel* positif yang diberi label dengan tepat oleh *classifier*, sementara *true negative* adalah *tupel* positif yang diberi label dengan tepat oleh *classifier*. *False positive* adalah *tupel* negatif yang diberi label tidak tepat, *false negative* adalah *tupel* positif yang diberi label dengan tidak tepat. Secara umum istilah ini berguna untuk menganalisis kemampuan *classifier* dan diringkas seperti tabel berikut.

Tabel 5. Model *Confusion Matrix*

	C1	C2
C	<i>True Positive</i>	<i>False Negative</i>
C	<i>False Positive</i>	<i>True Negative</i>

Dalam analisis pengenalan citra bunga ini variabel yang dihasilkan lebih dari 2 kelas, maka indikator untuk mendapatkan akurasi sebuah algoritma adalah menggunakan *precision*, *recall*, *accuracy*, dan *specificity*. Untuk penjelasan masing-masing indikator adalah sebagai berikut:

#### 3.6.1. Accuracy

Akurasi merupakan nilai persentase dari total seluruh data yang diidentifikasi dan dinilai benar. Persamaan untuk menghitung akurasi adalah:

$$Akurasi = True\ Positif / Total\ Dataset \tag{9}$$

$$Accuracy = \frac{True\ Positive}{\sum Dataset} \times 100\% \tag{10}$$

Di mana TP (*True Positive*) merupakan sebuah Kelas yang mampu diprediksi dengan benar dan tepat. Fungsi *True Positive* digunakan untuk menghitung nilai akurasi.

**3.6.2. Presisi**

Presisi merupakan data yang diambil berdasarkan informasi yang salah atau kurang. Persamaan untuk menghitung akurasi adalah:

$$Precision = \frac{True\ Positif}{True\ Positif + False\ Positif} \tag{11}$$

Setelah mendapatkan nilai *precision* dari masing-masing kelas, selanjutnya menghitung nilai *precision* dari data semua kelas, menggunakan persamaan sebagai berikut:

$$All\ Precision = \frac{P1+\dots+Pn}{\sum Variable} \times 100\% \tag{12}$$

**3.6.3. Recall**

*Recall* merupakan data yang tidak mampu diprediksi dengan benar. Persamaan untuk menghitung *recall* adalah:

$$Recall = \frac{True\ Positif}{True\ Positif + False\ Negative} \tag{13}$$

Setelah mendapatkan nilai *recall* dari masing-masing kelas, selanjutnya menghitung nilai *recall* dari data semua kelas, menggunakan persamaan sebagai berikut:

$$All\ Recall = \frac{R1+\dots+Rn}{\sum Variable} \times 100\% \tag{14}$$

**3.6.4. Specificity**

*Specificity* merupakan sebuah kelas yang mampu diprediksi dengan salah atau tidak tepat. Persamaan untuk menghitung *specificity* adalah:

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \tag{15}$$

Setelah mendapatkan nilai *specificity* dari masing-masing kelas, selanjutnya menghitung nilai *specificity* dari data semua kelas, menggunakan persamaan sebagai berikut:

$$All\ Specificity = \frac{S1+\dots+Sn}{\sum Variable} \times 100\% \tag{16}$$

Tabel 6. Hasil statistik data keseluruhan.

	Brodiaea elegans Hoover	Calochortus kennedyi Porter	Calochortus luteus Douglas ex	Calochortus splendens Douglas ex	Dichelostemma congestum (Sm.) Kunth	Encelia californica Nutt.	Hemerocallis fulva (L.) L.	Hypoxis hirsuta (L.) Coville	Lilium columbianum Leichtlin	Lilium pardalinum Kellogg	Lilium parvum Kellogg	Narcissus pseudonarcissus L.	Pyrrhopappus carolinianus (Walter) DC.	Tetrameuris scaposa (DC.) Greene	Tragopogon dubius Scop.	Tragopogon portifolius L.	Trillium erectum L.	Tussilago farfara L.	Wyethia angustifolia (DC.) Nutt.	
Brodiaea elegans Hoover	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Calochortus kennedyi Porter	0	12	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
Calochortus luteus Douglas ex	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Calochortus splendens Douglas ex	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dichelostemma congestum (Sm.) Kunth	2	0	0	1	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Encelia californica Nutt.	0	0	1	0	0	9	0	2	0	0	0	3	0	0	3	0	0	0	0	0
Hemerocallis fulva (L.) L.	0	0	0	0	0	2	7	0	0	2	3	1	0	0	0	0	0	0	0	0
Hypoxis hirsuta (L.) Coville	0	0	0	0	0	0	0	8	0	0	0	0	3	1	0	0	0	0	1	2
Lilium columbianum Leichtlin	0	0	0	0	0	0	0	0	7	1	0	0	0	0	0	0	0	0	0	0

Lilium pardalinum Kellogg	0	0	0	0	0	0	1	0	0	8	2	0	0	0	0	0	0	3	0	0
Lilium parvum Kellogg	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
Narcissus pseudonarcissus L.	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
Pyrrhopappus carolinianus (Walter) DC.	0	0	0	0	0	1	1	2	0	0	1	7	1	2	0	0	0	0	0	0
Tetraneuris scaposa (DC.) Greene	0	0	0	0	0	0	0	0	4	0	0	0	2	10	0	0	0	0	0	2
Tragopogon dubius Scop.	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0
Tragopogon porrifolius L.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0
Trillium erectum L.	1	0	0	2	1	0	3	0	0	0	0	0	0	0	3	9	0	0	0	0
Tussilago farfara L.	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0
Wyethia angustifolia (DC.) Nutt.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	8	0

Tabel 7. Hasil eksperimen.

Indikator	Persentase
Akurasi = TP / Total Dataset	71,1%
Precision = TP/(TP+FP)	71,1%
Recall = TP/(TP+FN)	77,8%
All Specificity = (Jumlah Specificity / Jumlah Kelas)	98,4%

Dari hasil uji coba menggunakan 228 gambar bunga dari data tes, dengan menggunakan metode Confusion Matrix data disimpulkan bahwa tingkat akurasi dengan menggunakan metode Naïve Bayes dari data ekstraksi warna dan bentuk dari citra bunga sebesar 71,1 %.

### 3.7. Area Under Curve

*Area Under Curve* atau biasa disingkat AUC digunakan untuk mengukur perbedaan performansi dalam sebuah algoritma. Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok [12]:

- a. 0.90 – 1.00 = Klasifikasi sangat baik
- b. 0.80 – 0.90 = Klasifikasi baik
- c. 0.70 – 0.80 = Klasifikasi cukup
- d. 0.60 – 0.70 = Klasifikasi buruk
- e. 0.50 – 0.60 = Klasifikasi salah

Pada penelitian ini untuk menghitung nilai AUC dengan menggunakan lebih dari dua kelas maka persamaan yang digunakan adalah:

$$AUC = \frac{Recall + Specificity}{2} \tag{17}$$

Maka dalam perhitungan yang dilakukan menggunakan citra bunga dari *dataset* maka nilai AUC adalah  $(77,8\% + 98,4\%) / 2 = 88\%$ . Dari hasil ini didapati bahwa analisis citra bunga menggunakan hasil Ekstraksi warna dan bentuk dengan metode Naïve Bayes termasuk dalam algoritma klasifikasi yang baik, sehingga data hasil klasifikasi Naïve Bayes sesuai dengan data citra sebenarnya.

## 4. Kesimpulan

Dari hasil penelitian yang dilakukan pada citra bunga berdasarkan warna dan bentuk menggunakan metode Naïve Bayes dapat diambil kesimpulan bahwa metode ini termasuk dalam algoritma yang baik untuk diaplikasikan pada klasifikasi citra bunga berdasarkan warna dan bentuk dengan nilai AUC sebesar 88% dan metode ini menghasilkan nilai akurasi sebesar 71,1%, sehingga pengenalan objek bunga menggunakan data hasil ekstraksi warna dan bentuk serta menggunakan metode klasifikasi Naïve Bayes sesuai dengan data citra sebenarnya

### Ucapan Terima Kasih

Peneliti mengucapkan terima kasih kepada LPPM Unpam dan Direktorat Riset dan Pengabdian Masyarakat Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset dan Teknologi/Badan Riset dan Inovasi Nasional Nomor:0105/D5/KP/LPPM/UNPAM/IV/2020 yang telah mendukung penelitian kami ini.

### Daftar Pustaka

- [1] dkk Nur Ridha Apriyanti, “Algoritma K-Means Clustering Dalam Pengolahan Citra Digital Landsat,” *Ilmu Komput.*, vol. 02, no. Clustering K-Measn, pp. 1–13, 2015.
- [2] P. Rosyani, “Deteksi Citra Bunga Menggunakan Analisis Segmentasi Fuzzy C-Means dan Otsu

- 
- Threshold,” vol. 20, no. 1, 2020.
- [3] P. Rosyani, “Pengenalan Wajah Menggunakan Metode Principal Component Analysis (PCA) dan Canberra Distance,” *J. Inform. Univ. Pamulang*, vol. 2, no. 2, p. 118, 2017.
- [4] Atina, “Segmentasi Citra Paru Menggunakan Metode k-Means Clustering,” *Segmentasi Citra Paru Menggunakan Metod. k-Means Clust.*, vol. 3, no. 2, pp. 57–65, 2017.
- [5] P. Rosyani, M. Taufik, A. A. Waskita, and D. H. Apriyanti, “Comparison of color model for flower recognition,” *2018 3rd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng.*, pp. 10–14, 2019.
- [6] D. H. Apriyanti, A. M. Arymurthy, and L. T. Handoko, “Identification of Orchid Species Using,” pp. 53–57, 2013.
- [7] I. Setiawan, W. Dewanta, H. A. Nugroho, and H. Supriyono, “Pengolah Citra Dengan Metode Thresholding Dengan Matlab R2014A,” *J. Media Infotama*, vol. 15, no. 2, pp. 65–70, 2019.
- [8] D. Rohpandi, A. Sugiharto, and M. Y. S. Jati, “Klasifikasi Citra Digital Berbasis Ekstraksi Ciri Berdasarkan Tekstur Menggunakan GLCM Dengan Algoritma K-Nearest Neighbor,” *J. Inform.*, vol. 3, no. 2, pp. 79–86, 2017.
- [9] A. A. Arifiyanti, R. M. Pradana, and I. F. Novian, “Klasifikasi Produk Retur dengan Algoritma Pohon Keputusan C4.5,” *J. IPTEK*, vol. 22, no. 1, p. 79, 2018.
- [10] E. Pudjiarti, “Prediksi Spam Email Menggunakan Metode Support Vector Machine Dan Particle Swarm Optimization,” *J. Pilar Nusa Mandiri*, vol. 12, no. 2, pp. 171–181, 2016.