

# IMPLEMENTASI METODE PRINCIPAL COMPONENT ANALYSIS (PCA) DAN MODIFIED K-NEAREST NEIGHBOR PADA KLASIFIKASI CITRA DAUN TANAMAN HERBAL

Nurdiansyah<sup>1</sup>, Muliadi<sup>\*2</sup>, Rudy Herteno<sup>3</sup>, Dwi Kartini<sup>4</sup>, Irwan Budiman<sup>5</sup>

<sup>1,2,3,4,5</sup> Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Lambung Mangkurat Banjarbaru  
*muliadi@ulm.ac.id*

## ABSTRAK

Klasifikasi citra tanaman herbal dapat dilakukan berdasarkan bentuk daun yang dikenal juga sebagai pengenalan citra tanaman herbal. Pada pengenalan citra dilakukan dengan mengidentifikasi ciri bentuk daun tanaman herbal dan dilakukan klasifikasi citra daun tersebut. Jumlah data citra yang digunakan sebagai 200 data yang terbagi kedalam 5 kelas. Sehingga masing-masing kelas terdiri dari 40 data citra. Masyarakat umumnya akan kesulitan untuk mengenal jenis tanaman herbal berdasarkan melihat secara sekilas pada daun. Pada metode PCA digunakan untuk mengurangi dimensi citra daun herbal dan metode Modified KNN digunakan untuk mengklasifikasikan citra daun herbal berdasarkan fitur. Pada pembagian data 6:4 nilai akurasi tertingginya sebesar 89 % pada K=2 dan akurasi terendahnya pada 78 % pada K=9. Pembagian data 7:3 nilai akurasi tertingginya sebesar 87 % pada K=4 dan akurasi terendahnya pada 82 % pada K=9. Dan Pada pembagian data 8:2 nilai akurasi tertingginya sebesar 93 % pada K=3 dan akurasi terendah sebesar 84 % pada K=4.

**Keyword :** *Tanaman Herbal, Principal Component Analysis (PCA), Modified K-Nearest Neighbor.*

## 1. PENDAHULUAN

Klasifikasi citra tanaman herbal dapat dilakukan berdasarkan bentuk daun yang dikenal juga sebagai pengenalan citra tanaman herbal. Pada pengenalan citra dilakukan dengan mengidentifikasi ciri bentuk daun tanaman herbal dan dilakukan klasifikasi citra daun tersebut. Berbagai teknik pengolahan citra telah dikembangkan untuk mempermudah pekerjaan manusia. Masyarakat umumnya akan kesulitan untuk mengenal jenis tanaman herbal berdasarkan melihat secara sekilas pada daun, Sebenarnya jika diamati dengan seksama akan bisa membedakan antara daun tanaman herbal satu dengan yang lain. Perlu adanya sistem pengolahan citra digital untuk membantu dalam mengenali tanaman herbal sehingga mempermudah dalam identifikasi dan klasifikasi tanaman herbal pada citra daun.

Pada penelitian yang dilakukan oleh [1] melakukan identifikasi kualitas buah tomat dengan menggunakan metode *Principal Component Analysis* (PCA) yang mana pada metode PCA ini mampu untuk mengurangi dimensi dataset dengan memproyeksikan setiap titik data ke hanya beberapa komponen utama untuk mendapatkan data berdimensi lebih rendah. Berdasarkan penelitian ini pula didapatkan hasil identifikasi kualitas data mencapai 76,7%. Nilai akurasi tersebut menunjukkan bahwa sistem dengan menggunakan metode PCA dapat berjalan dengan baik.

Kelebihan utama dari metode *Principal Component Analysis* (PCA) adalah metode yang

mengubah dimensi data yang banyak menjadi dimensi data yang sederhana dengan mempertahankan informasi dari dataset asli. Perbandingan metode *Principal Component Analysis* (PCA) berdasarkan *Linear Discriminant Analysis* (LDA) pada deteksi kanker. metode *Principal Component Analysis* (PCA) meningkatkan akurasi hingga 29,04 % dan f-1 score 64,28 % pada tumor usus dibandingkan tanpa menggunakan metode *Principal Component Analysis* (PCA) [2].

Setelah proses ekstraksi fitur citra, klasifikasi pada citra dapat dilakukan. *Modified K-Nearest Neighbor* merupakan pengembangan dari metode K-NN yang bertujuan untuk mengatasi masalah tingkat akurasi yang rendah pada algoritma K-NN. Algoritma MK-NN menambahkan dua prosedur dalam melakukan klasifikasi yaitu proses validasi dari data training serta proses pembobotan terbesar (weight voting) dari tetangga terdekatnya (nearest neighbor). Validasi dan bobot terbesar yang dimiliki oleh MK-NN mampu mengatasi kelemahan dari klasifikasi berdasarkan jarak terdekat sehingga metode MK-NN jauh lebih unggul dibandingkan metode algoritma K-NN [3]. Klasifikasi status gunung berapi menggunakan metode *Modified K-Nearest Neighbor* menghasilkan akurasi sebesar 86,67 % [4].

Hal ini dilakukan untuk mengetahui kinerja yang dihasilkan oleh metode *Principal Component Analysis* (PCA) dan *Modified K-Nearest Neighbor* dalam klasifikasi tanaman herbal pada citra daun.

## 2. TINJAUAN PUSTAKA

### 2.1. Literatur Terdahulu

Memprediksi penilaian akademis siswa menggunakan *Modified K Nearest Neighbor* (MKNN). Peneliti disini menggunakan algoritma *Modified K Nearest Neighbor* (MKNN) untuk mengkalkulasi jarak nilai tanpa memberikan kode. Pada penelitian menggunakan dataset penilaian siswa dengan 480 data siswa dengan 16 fitur [5].

Perbandingan tanaman jenis rimpang menggunakan *principal component analysis support vector machine*, *K-Nearest Neighbor* dan *Decision Tree*. Pada penelitian ini menggunakan terdapat 5 kelas tanaman rimpang yaitu jahe, kencur, kunyit, kunci dan laos. Dari hasil penelitian ini menghasilkan 98,2% pada SVM (Linear SVM), 90,02% pada KNN dan 87,3% pada Decision Tree [6].

Mendeteksi bibit jagung yang rusak karena pendingin. Pada penelitian ini menggunakan *Principal component analysis* (PCA) ekstraksi data dan menggunakan *K Nearest Neighbor* (KNN) sebagai metode untuk klasifikasinya. Pada penelitian ini diambil langsung pada suhu 13°C [7].

### 2.2. Tanaman Herbal

Tanaman herbal adalah tanaman dengan berbagai manfaat, salah satunya dapat digunakan untuk mengobati penyakit secara alami, terutama penyakit kulit dan penyakit rambut. Masyarakat Indonesia mudah terserang penyakit kulit dan rambut karena Indonesia adalah negara dengan iklim tropis. Di era modern ini sebagian besar masyarakat belum mahir untuk membedakan antara tanaman herbal dengan tanaman biasa sehingga dapat menyebabkan kesalahan dalam memilih tanaman herbal [8].

### 2.3. Pengolahan Citra

Pengolahan Citra Digital (Digital Image Processing) merupakan bidang ilmu yang mempelajari tentang bagaimana suatu citra itu dibentuk, diolah, dan dianalisis sehingga menghasilkan informasi yang dapat dipahami oleh manusia [9].

### 2.4. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) adalah suatu teknik statistik multivariat yang secara linear mengubah bentuk sekumpulan variabel asli menjadi kumpulan variabel yang lebih kecil yang tidak berkorelasi yang dapat mewakili informasi dari kumpulan variabel asli. Tujuan utamanya ialah menjelaskan sebanyak mungkin jumlah varian data asli dengan sedikit mungkin komponen utama yang disebut faktor [10].

### 2.5. Modified K-Nearest Neighbor

Metode *Modified K-Nearest Neighbor*, (MKNN) adalah Suatu algoritma yang dikembangkan dari algoritma K-Nearest Neighbor dengan menambahkan proses baru yaitu proses validitas data pada setiap dataset yang telah dilakukan perhitungan jarak euclidean antar pada data training, dan juga penambahan proses weight voting yang didapatkan dari hasil nilai validitas data dikalikan dengan hasil euclidean distance data training dan data testing [11].

### 2.6. Dataset

Dataset yang digunakan pada penelitian ini yaitu dataset citra daun tanaman herbal. Dataset ini memiliki total sebanyak 200 citra. Kemudian ada 5 kelas yaitu kelas daun Jambu, kunyit, lidah buaya, pepaya dan sirih. setiap kelas data memiliki 40 data citra daun. Dataset diabetes dibagi menjadi 3 pembagian data yaitu pembagian data 60% data training banding 40% data testing, 70% data training banding 30% data testing dan 80% data training banding 20% data testing.

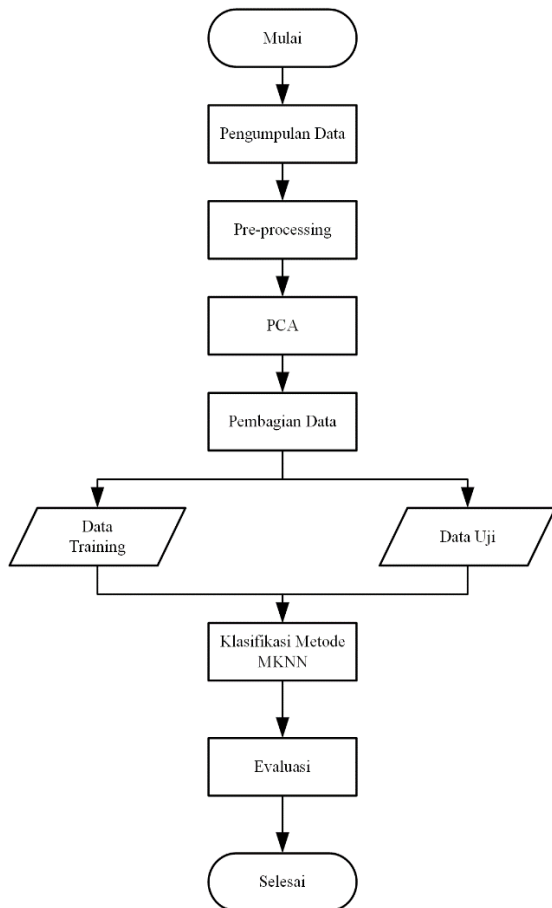
## 3. METODE PENELITIAN

Pada penelitian ini alur yang dilakukan dapat dilihat pada Gambar 1. Pada alur penelitian tersebut dapat diketahui langkah penelitian yang dilakukan dimulai dari pengumpulan data yaitu data citra daun meliputi daun jambu biji, kunyit, pepaya, sirih dan lidah buaya yang masing-masing berjumlah 40 data.

Kemudian dilanjutkan dengan tahapan preprocessing yang digunakan untuk membuang data yang tidak diperlukan yang kemudian dilanjutkan dengan mengubah ukuran citra (*resize*) dengan mengubah ukuran pixel. Setelah ukuran berubah dilanjutkan dengan mengubah citra tersebut kedalam bentuk *grayscale*.

Setelah proses preprocessing selesai dilakukan dilakukan ekstraksi fitur menggunakan *Principal Component Analysis* (PCA) yang digunakan untuk proses ekstraksi ciri pada data citra daun. Setelah didapatkan hasil ekstraksi dilanjutkan dengan pembagian data. Pada pembagian data dibagi menjadi dua yaitu data *training* dan *testing* dengan tiga perbandingan yang akan digunakan yaitu 60:40, 70:30, 80:20.

Setelah dilakukan pembagian data dilanjutkan dengan proses klasifikasi menggunakan Modified K-NN untuk melakukan proses identifikasi data citra daun tanaman. Kemudian dilanjutkan dengan proses evaluasi yang direpresentasikan menggunakan *confusion matrix* terhadap metode *Principal Component Analysis* (PCA) dan *Modified K-Nearest Neighbor* untuk mencari nilai *Precision*, *Recall*, *F1-score* dan Akurasi.



Gambar 1. Alur penelitian

### 3.1 Principal Component Analysis (PCA)

Metode PCA akan membentuk sekumpulan dimensi baru yang kemudian di ranking berdasarkan varian datanya. PCA akan menghasilkan Principal component yang didapat dari dekomposisi eigen value dan eigen vektor dari matriks kovariansi. Langkah dari algoritma PCA adalah sebagai berikut [12].

1. Perhitungan rata-rata (*mean*) :

Menghitung *mean* (X) dari data pada tiap dimensi menggunakan persamaan (1):

$$X = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

Dengan :

n = jumlah data sampel

X<sub>i</sub> = data sampel

2. Perhitungan matriks kovarian C

Menghitung *covariance matrix* (Cx) menggunakan persamaan (2)

$$C_x = \frac{1}{n-1} \sum_{i=1}^n (X_i - X)(X_i - X)^T \quad (2)$$

Dengan :

n = jumlah data sampel

X<sub>i</sub> = data sampel

3. Perhitungan eigenvalue dan eigenvector

Menghitung eigenvector (v<sub>m</sub>) dan eigenvalue(λ<sub>m</sub>) dari covarince matrix menggunakan persamaan (3):

$$C_x v_m = \lambda_m v_m \quad (3)$$

Dengan :

λ<sub>m</sub> = eigenvalue

v<sub>m</sub> = eigenvector

C<sub>x</sub> = matriks kovarian

4. Urutkan *eigenvalue* secara *descending*. *Principal Component* (PC) adalah deretan *eigenvector* sesuai dengan urutan *eigenvalue* pada tahap 3.
5. Menghasilkan *dataset* baru.

### 3.2 Modified K-Nearest Neighbor

Digunakan untuk melakukan pengklasifikasian untuk identifikasi kelas pada tanaman herbal. Dimana pada tahap klasifikasi data testing diuji dengan data training dilanjutkan dengan perhitungan validitas dan *weight voting*. Langkah-langkah pada metode Modified K-Nearest Neighbor sebagai berikut: [13]

#### 3.2.1 Perhitungan Jarak Euclidian

Untuk menghitung jarak Antara data training (x) dan data testing yang saling berdekatan dengan rumus sebagai berikut.

$$d(x_i, y_i) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (4)$$

p = Dimensi Data

d(x<sub>i</sub>, x<sub>y</sub>) = jarak dari data training dan testing

x<sub>i</sub> = data training

x<sub>y</sub> = data testing

#### 3.2.2 Perhitungan Nilai Validitas

Perhitungan nilai validitas dilakukan pada semua data training dengan rumus sebagai berikut.

$$\frac{1}{H} \sum_{i=0}^n S(lbl(x), lbl(Ni(x))) \quad (5)$$

H= Jumlah tetangga terdekat

lbl(x)= kelas data training

lbl(Ni(x))= label kelas titik terdekat data training

Fungsi S untuk menghitung kesamaan antara titik x dan data ke-i dari tetangga terdekat.

$$S(a,b) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases} \quad (6)$$

#### 3.2.3 Perhitungan Weighted Voting

Perhitungan *weight voting* memiliki pengaruh penting terhadap data nilai validitas dan nilai jarak terdekat. Dengan perhitungan *weight voting* membantu dalam mengatasi kelemahan pada metode KNN yang didasarkan hanya pada perhitungan nilai jarak antar data. Berikut rumus persamaan *weight voting* sebagai berikut :

$$W(i) = \text{Validity}(i) \times \frac{1}{de+0.5} \quad (7)$$

W(i) = Weighted Voting

Validity(i)= nilai validitas

De = jarak Euclidean

**3.3. Confusion matrix**

Alat ukur performa klasifikasi dari kedua metode yang digunakan adalah *confusion matrix* dengan dengan TP (True Positive) dan TF (True Negatif) sebagai variabel yang memprediksi benar sedangkan TN (True Negative) dan (False Positive) sebagai variabel yang memprediksi salah.

Tabel 1. Confusion Matriks

Kelas Sebenarnya	POSITIVE	NEGATIVE
POSITIVE	TP	TN
NEGATIVE	FP	FN

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} * 100\% \quad (8)$$

$$Precision = \frac{TP}{TP+FP} * 100\% \quad (9)$$

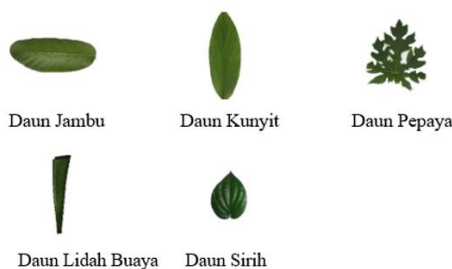
$$Recall = \frac{TP}{TP+FN} * 100\% \quad (10)$$

$$F1\ Score = \frac{2 * (Recall * Precision)}{Recall + Precision} * 100\% \quad (11)$$

Tahapan penelitian yang dilaksanakan dalam penelitian ini yaitu Pengumpulan data citra, *Preprocessing Data*, Ekstraksi Fitur *Principal Component Analysis*, Pembagian Data, Klasifikasi Modified K-NN, dan Evaluasi yang direpresentasikan menggunakan *confusion matrix* terhadap metode *Principal Component Analysis* (PCA) dan *Modified K-Nearest Neighbor* yang digunakan dan nilai yang dicari adalah nilai *Precision*, *Recall*, *F1-score* dan Akurasi.

**4. HASIL DAN PEMBAHASAN**

Pada penelitian ini, menggunakan data citra tanaman dengan 5 kelas data yaitu Jambu, Kunyit, Lidah Buaya, pepaya dan Sirih yang dapat dilihat pada Gambar 2 berikut.

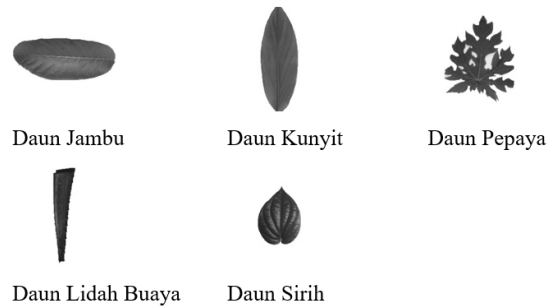


Gambar 2. Data citra daun

Tahapan ekstraksi fitur dilakukan dengan menggunakan metode *Principal Component Analysis* (PCA). Selanjutnya setelah diekstraksi fitur, fitur *numeric* yang dihasilkan dari tahap ekstraksi fitur *Principal Component Analysis* (PCA) diklasifikasi dengan metode klasifikasi *Modified K-Nearest Neighbor* dengan nilai k yang bervariasi yaitu K=1 sampai dengan K=10 menggunakan split data dengan pembagian 6 : 4, 7 : 3 dan 8 : 2. Pengujian ini dimaksudkan untuk mengetahui

bagaimana kinerja dari kombinasi metode *Principal Component Analysis* (PCA) dan *Modified K-Nearest Neighbor* serta untuk mengetahui akurasi tertinggi yang didapatkan dari penelitian ini. Setelah semua data diujikan selanjutnya hasil pengujian akan dievaluasi menggunakan *Confusion Matrix* untuk memperoleh nilai akurasi, *precision*, *recall*, dan *F1-score* pada setiap nilai K.

Adapun perubahan data citra setelah melalui proses *preprocessing* yaitu membuat gambar menjadi *grayscale* dapat dilihat pada Gambar 3 berikut.



Gambar 3. Data citra daun dalam grayscale

**4.1. Data Kelas Jambu**

Tabel 2. Hasil *Precision* pada kelas Jambu Biji

Nilai K	Precision Jambu Biji		
	6 : 4	7 : 3	8 : 2
1	76 %	67 %	80 %
2	88 %	79 %	100 %
3	75 %	73 %	100 %
4	79 %	79 %	100 %
5	83 %	85 %	88 %
6	79 %	85 %	100 %
7	75 %	92 %	100 %
8	79 %	85 %	88 %
9	75 %	85 %	88 %
10	75 %	85 %	88 %

Pada kelas Jambu biji didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan semakin tinggi nilai K maka nilai *precision* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang bukan dari data sebenarnya. Pada pembagian data 7:3 menunjukkan semakin tinggi nilai K maka nilai *precision* cenderung menaik dikarenakan banyak data yang dibandingkan dengan data yang sama dari data sebenarnya. Dan pada pembagian data 8:2 menunjukkan semakin tinggi nilai K maka nilai *precision* cenderung menaik dikarenakan banyak data yang dibandingkan dengan data yang sama dari data sebenarnya. Hal ini menunjukkan bahwa pada pembagian data yang berbeda menghasilkan nilai *precision* yang berbeda.

Tabel 3. Hasil *Recall* Pada Kelas Jambu biji

Recall Jambu Biji			
Nilai k	6 : 4	7 : 3	8 : 2
1	100 %	100 %	100 %
2	94 %	92 %	88 %
3	94 %	92 %	88 %
4	94 %	92 %	88 %
5	94 %	92 %	88 %
6	94 %	92 %	88 %
7	94 %	92 %	88 %
8	94 %	92 %	88 %
9	94 %	92 %	88 %
10	94 %	92 %	88 %

Pada kelas Jambu biji didapatkan hasil pengujian pada pembagian data 6:4, pada pembagian data 7:3 dan pada pembagian data 8:2 menunjukkan hasil yang sama pada saat K= 2 maka nilai *recall* menurun dan stabil dikarenakan pada saat pengujian data yang memiliki kemiripan data yang cukup tinggi dengan data yang memiliki kelas berbeda, sehingga terjadi kesalahan klasifikasi pada sistem yang menyebabkan tingkat keberhasilan sistem dalam memprediksi data positif menurun.

Tabel 4. Hasil *F1-score* Pada kelas Jambu Biji

F1-score Jambu Biji			
Nilai k	6 : 4	7 : 3	8 : 2
1	86 %	80 %	89 %
2	91 %	85 %	93 %
3	83 %	81 %	93 %
4	86 %	85 %	93 %
5	88 %	88 %	88 %
6	86 %	88 %	93 %
7	83 %	92 %	93 %
8	86 %	88 %	88 %
9	83 %	88 %	88 %
10	83 %	88 %	88 %

Pada kelas Jambu biji didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan bahwa semakin tinggi *F1-Score* maka nilai K semakin rendah. Pada pembagian pembagian data 7:3 menunjukkan semakin tinggi nilai K maka semakin tinggi nilai *F1-score*. Dan pada pembagian data 8:2 menunjukkan bahwa nilai *F1-Score* semakin tinggi nilai K maka semakin tinggi nilai *F1-score* tetapi pada K=5 dan K=8 maka nilai *F1-score* turun. Hal ini menunjukkan bahwa perbandingan pada nilai *precision* dan *recall* semakin tinggi maka semakin bagus.

#### 4.2. Kelas Data Kunyit

Tabel 5. Tabel *Precision* pada Kelas Kunyit

Precision Kunyit			
Nilai K	6 : 4	7 : 3	8 : 2
1	85 %	75 %	75 %
2	82 %	75 %	73 %
3	86 %	80 %	80 %
4	81 %	82 %	80 %
5	82 %	77 %	78 %

Precision Kunyit			
Nilai K	6 : 4	7 : 3	8 : 2
6	76 %	69 %	73 %
7	65 %	71 %	73 %
8	65 %	67 %	73 %
9	67 %	67 %	73 %
10	67 %	71 %	80 %

Pada kelas Kunyit didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan semakin tinggi nilai K maka nilai *precision* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang bukan dari data sebenarnya. Pada pembagian data 7:3 menunjukkan semakin tinggi nilai K maka nilai *precision* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang sama dari data sebenarnya. Pembagian data 8:2 menunjukkan semakin tinggi nilai K maka nilai *precision* juga cenderung naik dikarenakan banyaknya data yang dibandingkan dengan data yang sama dari data sebenarnya.

Tabel 6. Tabel *Recall* pada Kelas Kunyit

Recall Kunyit			
Nilai K	6 : 4	7 : 3	8 : 2
1	69 %	50 %	75 %
2	88 %	75 %	100 %
3	75 %	67 %	100 %
4	81 %	75 %	100 %
5	88 %	83 %	88 %
6	81 %	75 %	100 %
7	69 %	83 %	100 %
8	81 %	83 %	100 %
9	75 %	83 %	100 %
10	75 %	83 %	100 %

Pada kelas Kunyit didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan semakin tinggi nilai K maka nilai *recall* cenderung naik dikarenakan pada saat pengujian data yang memiliki kemiripan cukup tinggi dengan data yang memiliki kelas berbeda. Pada pembagian data 7:3 menunjukkan semakin tinggi nilai K maka nilai *recall* cenderung menaik dikarenakan pada saat pengujian data yang memiliki kemiripan data yang cukup tinggi dengan data yang memiliki kelas berbeda. Kemudian pembagian data 8:2 menunjukkan semakin tinggi nilai K maka nilai *recall* cenderung menaik dikarenakan pada saat pengujian data yang memiliki kemiripan data yang cukup tinggi dengan data yang memiliki kelas berbeda. Hal ini menunjukkan bahwa pada pembagian data yang berbeda menghasilkan nilai *recall* yang sama.

Tabel 7. Tabel *F1-score* pada Kelas Kunyit

F1-score Kunyit			
Nilai K	6 : 4	7 : 3	8 : 2
1	76 %	60 %	75 %
2	85 %	75 %	84 %
3	80 %	73 %	89 %
4	81 %	78 %	89 %
5	85 %	80 %	82 %
6	79 %	72 %	84 %
7	67 %	77 %	84 %
8	72 %	74 %	84 %
9	71 %	74 %	84 %
10	71 %	77 %	89 %

Pada kelas Jambu biji didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan bahwa semakin tinggi nilai K maka semakin rendah nilai *F1-score*. Pada pembagian data 7:3 menunjukkan nilai *F1-Score* semakin tinggi nilai K maka semakin tinggi nilai *F1-score*. Pembagian data 8:2 menunjukkan semakin tinggi nilai K maka semakin tinggi nilai *F1-score*.

4.3. Kelas Data Lidah Buaya

Tabel 8. Tabel Precision pada Kelas Lidah Buaya

Precision Lidah Buaya			
Nilai K	6 : 4	7 : 3	8 : 2
1	100 %	100 %	100 %
2	100 %	100 %	100 %
3	100 %	100 %	100 %
4	100 %	100 %	100 %
5	100 %	100 %	100 %
6	100 %	100 %	100 %
7	100 %	100 %	100 %
8	100 %	100 %	100 %
9	100 %	100 %	100 %
10	100 %	100 %	100 %

Pada kelas Lidah Buaya didapatkan hasil pengujian pada pembagian data 6:4,7:3 dan 8:2 menunjukkan hasil yang sama pada dari k=1 sampai k = 10 maka nilai *precision* tinggi dan stabil dikarenakan banyak data yang dibandingkan dengan data yang sama dari data sebenarnya. Hal ini menunjukkan bahwa pada pembagian data yang berbeda menghasilkan nilai *precision* yang sama.

Tabel 9. Tabel Recall pada Kelas Lidah Buaya

Recall Lidah Buaya			
Nilai K	6 : 4	7 : 3	8 : 2
1	81 %	83 %	75 %
2	81 %	83 %	75 %
3	81 %	92 %	88 %
4	75 %	83 %	75 %
5	69 %	75 %	75 %
6	62 %	75 %	75 %
7	50 %	75 %	75 %
8	44 %	58 %	62 %
9	50 %	58 %	62 %
10	50 %	67 %	75 %

Pada kelas Lidah buaya didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan semakin tinggi nilai K maka *recall* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang bukan dari data sebenarnya. Pada pembagian data 7:3 menunjukkan semakin tinggi nilai K maka nilai *recall* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang sama dari data sebenarnya. Pembagian data 8:2 menunjukkan semakin tinggi nilai K maka nilai *recall* cenderung menurun dikarenakan pada saat pengujian data yang memiliki kemiripan data yang cukup tinggi dengan data yang memiliki kelas berbeda, sehingga terjadi kesalahan klasifikasi pada system yang menyebabkan tingkat keberhasilan memprediksi data positif menurun.

Tabel 10. Tabel *F1-score* pada Kelas Lidah Buaya

F1-score Lidah Buaya			
Nilai K	6 : 4	7 : 3	8 : 2
1	90 %	91 %	86 %
2	90 %	91 %	86 %
3	90 %	96 %	93 %
4	86 %	91 %	86 %
5	81 %	86 %	86 %
6	77 %	86 %	86 %
7	67 %	86 %	86 %
8	61 %	74 %	77 %
9	67 %	74 %	77 %
10	67 %	80 %	86 %

Pada kelas lidah buaya didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan semakin tinggi nilai K maka nilai *F1-score* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang bukan dari data sebenarnya. Pada pembagian data 7:3 menunjukkan semakin tinggi nilai K maka nilai *F1-score* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang sama dari data sebenarnya. Dan pada pembagian data 8:2 menunjukkan semakin tinggi nilai K maka nilai *F1-score* cenderung menurun. Hal ini menunjukkan bahwa perbandingan pada nilai *precision* dan *recall* semakin tinggi maka semakin kurang bagus.

4.4. Kelas Data Pepaya

Tabel 11. Tabel Precision pada Kelas Pepaya

Precision Pepaya			
Nilai K	6 : 4	7 : 3	8 : 2
1	93 %	100 %	100 %
2	93 %	100 %	100 %
3	93 %	100 %	100 %
4	93 %	91 %	83 %
5	88 %	91 %	88 %
6	93 %	100 %	100 %
7	86 %	100 %	100 %
8	92 %	100 %	100 %

Precision Pepaya			
Nilai K	6 : 4	7 : 3	8 : 2
9	92 %	100 %	100 %
10	92 %	100 %	100 %

Pada kelas pepaya didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan bahwa nilai *precision* semakin tinggi nilai K maka semakin rendah nilai *precision*. Pada pembagian data 7:3 menunjukkan bahwa nilai *precision* semakin tinggi nilai K maka semakin tinggi nilai *precision* dan pada K=4 sampai k=5 terjadi penurunan. Dan pada pembagian data 8:2 menunjukkan bahwa nilai *precision* semakin tinggi nilai K maka semakin tinggi nilai *precision* dan pada K=4 sampai k=5 terjadi penurunan. dikarenakan banyak data yang dibandingkan dengan data yang sama dari data sebenarnya. Hal ini menunjukkan bahwa pada pembagian data yang berbeda menghasilkan nilai *precision* yang berbeda.

Tabel 12. Tabel Recall pada Kelas Pepaya

Recall Pepaya			
Nilai K	6 : 4	7 : 3	80 : 20
1	81 %	92 %	88 %
2	81 %	75 %	62 %
3	88 %	83 %	88 %
4	81 %	83 %	62 %
5	88 %	83 %	88 %
6	81 %	75 %	75 %
7	75 %	83 %	88 %
8	69 %	75 %	88 %
9	69 %	75 %	75 %
10	69 %	75 %	75 %

Pada kelas pepaya didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan semakin tinggi nilai K maka nilai *recall* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang bukan dari data sebenarnya. Pada pembagian data 7:3 menunjukkan semakin tinggi nilai K maka nilai *recall* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang sama dari data sebenarnya. Dan pada pembagian data 8:2 menunjukkan semakin tinggi nilai K maka nilai *recall* cenderung menurun dikarenakan pada saat pengujian data yang memiliki kemiripan data yang cukup tinggi dengan data yang memiliki kelas berbeda.

Tabel 13. Tabel F1-score pada Kelas Pepaya

F1-score Pepaya			
Nilai K	6 : 4	7 : 3	8 : 2
1	87 %	96 %	93 %
2	87 %	86 %	77 %
3	90 %	91 %	93 %
4	87 %	87 %	71 %
5	88 %	87 %	88 %
6	87 %	86 %	86 %
7	80 %	91 %	93 %
8	79 %	86 %	93 %

F1-score Pepaya			
Nilai K	6 : 4	7 : 3	8 : 2
9	79 %	86 %	86 %
10	79 %	86 %	86 %

Pada kelas pepaya didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan semakin tinggi nilai K maka nilai *F1-score* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang bukan dari data sebenarnya. Pada pembagian data 7:3 menunjukkan semakin tinggi nilai K maka nilai *F1-score* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang sama dari data sebenarnya. Dan pada pembagian data 8:2 menunjukkan semakin tinggi nilai K maka nilai *F1-score* cenderung menurun. Hal ini menunjukkan bahwa perbandingan pada nilai *precision* dan *recall* semakin tinggi maka semakin kurang bagus.

#### 4.5. Kelas Data Sirih

Tabel 14. Tabel Precision pada Kelas Sirih

Precision Sirih			
Nilai K	6 : 4	7 : 3	8 : 2
1	84 %	92 %	89 %
2	84 %	80 %	73 %
3	89 %	86 %	89 %
4	84 %	86 %	73 %
5	89 %	86 %	89 %
6	80 %	75 %	80 %
7	76 %	80 %	89 %
8	73 %	75 %	89 %
9	73 %	75 %	80 %
10	73 %	75 %	80 %

Pada kelas sirih didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan nilai *precision* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang bukan dari data sebenarnya. Pada pembagian data 7:3 menunjukkan semakin tinggi nilai K maka nilai *precision* cenderung menurun dikarenakan banyak data yang dibandingkan dengan data yang sama dari data sebenarnya. Dan pada pembagian data 8:2 menunjukkan semakin tinggi nilai K maka nilai *precision* cenderung menurun.

Tabel 15. Tabel Recall pada Kelas Sirih

Recall Sirih			
Nilai K	6 : 4	7 : 3	8 : 2
1	100 %	100 %	100 %
2	100 %	100 %	100 %
3	100 %	100 %	100 %
4	100 %	100 %	100 %
5	100 %	100 %	100 %
6	100 %	100 %	100 %
7	100 %	100 %	100 %
8	100 %	100 %	100 %
9	100 %	100 %	100 %
10	100 %	100 %	100 %

Pada kelas Sirih didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan bahwa nilai *recall* semakin tinggi nilai K maka semakin rendah nilai *recall*. Pada pembagian data 7:3 menunjukkan bahwa nilai *recall* semakin tinggi nilai K maka semakin tinggi nilai *recall*. Dan pada pembagian data 8:2 menunjukkan bahwa nilai *recall* semakin tinggi nilai K maka semakin tinggi nilai *recall*. Hal ini dikarenakan pada saat pengujian data yang memiliki kemiripan data yang cukup tinggi dengan data yang memiliki kelas berbeda.

Tabel 16. Tabel F1-score pada Kelas Sirih

F1-score Sirih			
Nilai K	6 : 4	7 : 3	8 : 2
1	91 %	96 %	94 %
2	91 %	89 %	84 %
3	94 %	92 %	94 %
4	91 %	92 %	84 %
5	94 %	92 %	94 %
6	89 %	86 %	89 %
7	86 %	89 %	94 %
8	84 %	86 %	94 %
9	84 %	86 %	89 %
10	84 %	86 %	89 %

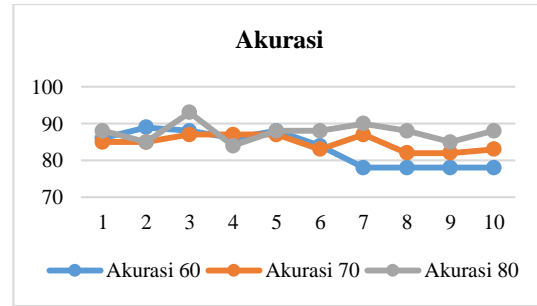
Pada kelas Sirih didapatkan hasil pengujian pada pembagian data 6:4 menunjukkan bahwa nilai *F1-score* semakin tinggi nilai K maka semakin rendah nilai *F1-score*. Pada pembagian data 7:3 menunjukkan bahwa nilai *F1-score* semakin tinggi nilai K maka semakin rendah nilai *F1-score*. Dan pada pembagian data 8:2 menunjukkan bahwa nilai *F1-score* semakin tinggi nilai K maka semakin rendah nilai *f1-score*. Hal ini menunjukkan bahwa perbandingan pada nilai *precision* dan *recall* semakin tinggi maka semakin kurang bagus.

4.6. Akurasi

Tabel 17. Tabel Akurasi

Akurasi (%)			
Nilai K	6 : 4	7 : 3	8 : 2
1	86 %	85 %	88 %
2	89 %	85 %	85 %
3	88 %	87 %	93 %
4	86 %	87 %	84 %
5	88 %	87 %	88 %
6	84 %	83 %	88 %
7	78 %	87 %	90 %
8	78 %	82 %	88 %
9	78 %	82 %	85 %
10	78 %	83 %	88 %

Pada gambar 1 menunjukkan bahwa akurasi didapatkan bahwa pada nilai K yang mempunyai nilai kecil maka nilai akurasi yang dihasilkan tinggi sedangkan pada saat nilai K semakin besar maka nilai akurasi yang dihasilkan semakin menurun dibandingkan nilai K yang kecil.



Gambar 1. Hasil Perhitungan Akurasi

5. KESIMPULAN DAN SARAN

Penelitian yang telah dilakukan dapat diambil kesimpulan bahwa penerapan metode Principal Component Analysis (PCA) Modified K-Nearest Neighbor dapat digunakan untuk mengidentifikasi dan mengklasifikasi jenis tanaman herbal khususnya pada daun tanaman jambu, kunyit, pepaya, lidah buaya dan sirih. Pada pembagian data 6:4 nilai akurasi tertingginya sebesar 89 % pada K=2 dan akurasi terendahnya pada 78 % pada K=9. Pembagian data 7:3 nilai akurasi tertingginya sebesar 87 % pada K=4 dan akurasi terendahnya pada 82 % pada K=9. Dan Pada pembagian data 8:4 nilai akurasi tertingginya sebesar 93 % pada K=3 dan akurasi terendah sebesar 84 % pada K=4.

Pada penelitian selanjutnya, disarankan untuk menambahkan metode klasifikasi yang berbeda agar mengetahui hasil nilai akurasi yang lebih bagus dan menggunakan k-fold cross validation pada pembagian data nya sehingga didapat nilai akurasi yang lebih tinggi.

DAFTAR PUSTAKA

- [1] U. Murdika, Muhammad Alif, Yessi Mulyani, "Identifikasi Kualitas Buah Tomat dengan Metode PCA (Principal Component Analysis) dan Backpropagation," *Electrician*, vol. 15, no. 3, pp. 175–180, 2021, doi: 10.23960/elc.v15n3.2240.
- [2] A. Astuti, W., & Adiwijaya, "Principal Component Analysis Sebagai Ekstraksi Fitur Data Microarray Untuk Deteksi Kanker Berbasis Linear Discriminant Analysis.," *J. Media Inform. Budidarma*, vol. 3, no. 2, pp. 72–77, 2019.
- [3] S. Amelia, "Penerapan Metode Modified K-Nearest Neighbor pada Pengklasifikasian Status Pembayaran Kredit Barang Elektronik dan Furniture," *Stat. J. Theor. Stat. Its Appl.*, vol. 22, no. 1, pp. 95–104, 2022, doi: 10.29313/statistika.v22i1.345.
- [4] M. T. Anggian, F. C., Hidayat, N., & Furqon, "Implementasi Metode Modified K-Nearest Neighbor untuk Klasifikasi Status Gunung Berapi.," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN, 2548, 964X.*, 2020.
- [5] V. Jawthari, M., & Stoffová, "Predicting students' academic performance using a



- modified kNN algorithm.," *Pollack Period.*, vol. 16, no. 3, pp. 20–26, 2021.
- [6] M. B. Mayasari, M., Mulyana, D. I., & Yel, "Komparasi Klasifikasi Jenis Tanaman Rimpang Menggunakan Principal Component Analisis, Support Vector Machine, K-Nearest Neighbor Dan Decision Tree.," *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 6, no. 2, pp. 644–655, 2022.
- [7] F. Zhang, J., Wang, Z., Qu, M., & Cheng, "Research on physicochemical properties, microscopic characterization and detection of different freezing-damaged corn seeds.," *Food Chem.*, vol. X, pp. 100–338, 2022.
- [8] A. M. Atha and E. Zuliarso, "Deteksi Tanaman Herbal Khusus Untuk Penyakit Kulit Dan Penyakit Rambut Menggunakan Convolutional Neural Network (CNN) Dan Tensorflow," *JUPITER (Jurnal Penelit. Ilmu dan Teknol. Komputer)*, vol. 14, no. 2-a, pp. 1–10, 2022.
- [9] R. Silvia, "Pengolahan citra digital dan histogram dengan phyton dan text editor phycharm," *Technol. J. Ilm.*, vol. 11, no. 3, pp. 181–186, 2020.
- [10] M. Wangge, "Penerapan Metode Principal Component Analysis (PCA) Terhadap Faktor-faktor yang Mempengaruhi Lamanya Penyelesaian Skripsi Mahasiswa Program Studi Pendidikan Matematika FKIP UNDANA," *J. Cendekia J. Pendidik. Mat.*, vol. 5, no. 2, pp. 974–988, 2021, doi: 10.31004/cendekia.v5i2.465.
- [11] A. I. Saputra, H. Oktavianto, H. Azizah, and A. Faruq, "Penerapan Algoritma Modified K-Nearest Neighbour (MKNN) Pada Klasifikasi Masa Studi Mahasiswa Teknik Informatika Application of the Modified K-Nearest Neighbour (MKNN) Algorithm In the Classification of the Study Period of Informatics Engineering Student," *J. Smart Teknol.*, vol. 3, no. 1, pp. 2774–1702, 2021, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JS T>.
- [12] D. Hedyati and I. M. Suartana, "Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro," *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 2, pp. 49–54, 2021, doi: 10.26740/jieet.v5n2.p49-54.
- [13] S. A. Ravi, M. R., & Indriati, "Implementasi Algoritme Modified K-Nearest Neighbor (MKNN) Untuk Mengidentifikasi Penyakit Gigi Dan Mulut.," *J. Pengemb. Teknol. Inf. dan Ilmu Komputere-ISSN*, 2596-2602., 2019.