

1

Identifikasi Jenis Kayu Menggunakan Support Vector Machine Berbasis Data Citra

A.A Gede Rai Gunawan, Sri Nurdiati, Yandra Akeman

9

Klasifikasi Metagenom dengan Metode *Naïve Bayes Classifier*

Dian Kartika Utami, Wisnu Ananta Kusuma, Agus Buono

18

Identifikasi Daun *Shorea* menggunakan KNN dengan Ekstraksi Fitur 2DPCA

Erni Yusniar, Aziz Kustiyo

27

Pengenalan Suara Paru-Paru dengan MFCC sebagai Ekstraksi Ciri dan *Backpropagation* sebagai *Classifier*

Fadhilah Syafria, Agus Buono, Bib Paruhum Silalahi

37

Optimasi *K-Means Clustering* Menggunakan *Particle Swarm Optimization* pada Sistem Identifikasi Tumbuhan Obat Berbasis Citra

Franki Yusuf Bisilisin, Yeni Herdiyeni, Bib Paruhum Silalahi

47

Sistem Informasi Geografis Persebaran Titik Api di Indonesia Menggunakan OpenGeo Suite 3.0

Sonita Veronica Br Barus, Imas Sukaesih Sitanggang

57

Pengembangan dan Uji Usability Sistem Informasi Manajemen Pemantauan Kehadiran dan Nilai Ujian Siswa

Suskamiyadi, Yani Nurhadryani, Heru Sukoco



## Optimasi *K-Means Clustering* Menggunakan *Particle Swarm Optimization* pada Sistem Identifikasi Tumbuhan Obat Berbasis Citra

### *K-Means Clustering Optimization Using Particle Swarm Optimization on Image Based Medicinal Plant Identification System*

FRANKI YUSUF BISILISIN<sup>1</sup>, YENI HERDIYENI<sup>1\*</sup>, BIB PARUHUM SILALAH<sup>2</sup>

#### Abstrak

Teknologi identifikasi pada penelitian ini diperlukan untuk mempercepat proses identifikasi spesies tumbuhan obat berupa data citra *digital*. Penelitian ini membangun sistem identifikasi tumbuhan obat menggunakan teknik *clustering*. Teknik *clustering* digunakan untuk mengelompokkan data citra sesuai dengan spesies tumbuhan obat. Penelitian ini bertujuan melakukan optimasi *k-means clustering* menggunakan metode *particle swarm optimization* (PSO). Metode PSO digunakan untuk mengatasi kelemahan pada metode *clustering* tradisional yaitu pemilihan pusat *cluster* awal dan solusi lokal. Proses ekstraksi fitur menggunakan *fuzzy local binary pattern* (FLBP) untuk merepresentasikan tekstur dari citra. Implementasi program menggunakan bahasa pemrograman C++. Analisis *clustering* dilakukan untuk 30 spesies tumbuhan obat yang ada di Indonesia dengan jumlah 48 citra masing-masing spesies. Pengukuran kualitas *clustering* menggunakan nilai *quantization error* dan akurasi. Hasil yang diperoleh menunjukkan metode PSO mampu meningkatkan kinerja dari metode *k-means clustering* dalam proses identifikasi tumbuhan obat.

Kata kunci: *fuzzy local binary pattern, k-means clustering, particle swarm optimization, tumbuhan obat*

#### Abstract

*Automatic identification technology is required to accelerate the process of identification of medicinal plant species in the form of digital image data. This research used clustering techniques to build the system identification of medicinal plants. Clustering technique was used to classify the image data according to the species of medicinal plants. The purpose of this study was to optimize the k-means clustering using particle swarm optimization (PSO). PSO method was used to overcome the shortcomings of traditional clustering algorithms such as the selection of the initial cluster centers and local solutions. The feature extraction was conducted by using fuzzy local binary pattern (FLBP) to represent the texture of the image. The program was implemented in C++ programming language. Clustering analysis was performed for 30 species of Indonesian medicinal plants with 48 images of each species. The values of quantization error and accuracy were utilized to measure the quality of clustering. Clustering results showed that PSO was able to improve the performance of k-means clustering in the identification process of medicinal plants.*

Keywords: *fuzzy local binary pattern, k-means clustering, medicinal plant, particle swarm optimization*

## PENDAHULUAN

Indonesia merupakan negara dengan tingkat keanekaragaman hayati yang tinggi. Groombridge dan Jenkins (2002) mencatat bahwa terdapat 22 500 spesies tumbuhan obat di Indonesia. Jumlah spesies yang telah dimanfaatkan sebagai tumbuhan obat sebanyak 1000

<sup>1</sup>Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Bogor 16680

<sup>2</sup>Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Bogor 16680

\*Penulis korespondensi: Tel/Faks: 0251-8625584; Surel: [yeni\\_herdiyeni@ipb.ac.id](mailto:yeni_herdiyeni@ipb.ac.id)

spesies. Hal ini menunjukkan bahwa jumlah persentase tumbuhan obat yang telah dimanfaatkan hanya sebesar 4.4% dari sumber daya tumbuhan obat yang tersedia. Salah satu penyebab kurangnya pemanfaatan tumbuhan obat adalah minimnya pengetahuan masyarakat mengenai potensi tumbuhan obat. Peningkatan pengetahuan masyarakat tentang tumbuhan obat dapat dibantu dengan dikembangkannya sistem identifikasi tumbuhan obat.

Teknologi identifikasi secara otomatis diperlukan untuk mempercepat proses identifikasi dengan menggunakan organ vegetatif seperti daun. Sistem identifikasi terdiri atas dua teknik, yaitu teknik klasifikasi dan teknik *clustering*. Adapun penelitian menggunakan teknik klasifikasi yang sudah dilakukan, salah satunya pada penelitian Herdiyeni dan Wahyuni (2012) yang menggunakan metode ekstraksi fitur *fuzzy local binary pattern* (FLBP) dan metode klasifikasi *probabilistic neural network* (PNN) untuk identifikasi tumbuhan obat. Dengan menggunakan teknik klasifikasi banyak data tumbuhan obat yang belum teridentifikasi sehingga teknik *unsupervised learning*, yakni *clustering*, diusulkan untuk membandingkan hasil identifikasi tumbuhan obat dengan teknik klasifikasi.

Menurut Jain *et al.* (1999), *clustering* adalah ilmu yang mempelajari hubungan antara satu data dan data yang lain kemudian mengelompokkannya menjadi satu kategori tertentu. Metode *clustering* yang paling banyak digunakan ialah metode *k-means clustering*. Kelemahan utama dari metode ini adalah hasil yang sensitif terhadap pemilihan pusat *cluster* awal dan perhitungan solusi lokal untuk mencapai kondisi optimal. Menurut Panchal *et al.* (2009), pemilihan pusat *cluster* awal dan perhitungan solusi lokal mempengaruhi proses dari kinerja metode dan hasil partisi data. Kelemahan pemilihan pusat *cluster* awal dan perhitungan solusi lokal pada metode *k-means clustering* memberikan motivasi untuk menerapkan metode *particle swarm optimization* (PSO). Menurut Panchal *et al.* (2009), PSO dapat digunakan untuk menemukan pusat *cluster* baru sesuai dengan jumlah *cluster* yang sudah ditentukan. Algoritme dimodifikasi menggunakan metode *k-means clustering* untuk pengelompokan awal, kemudian PSO memperbaiki kelompok data yang dibentuk oleh *k-means clustering*. Nilai *quantization error* dan akurasi digunakan untuk mengukur seberapa baik metode PSO yang diterapkan.

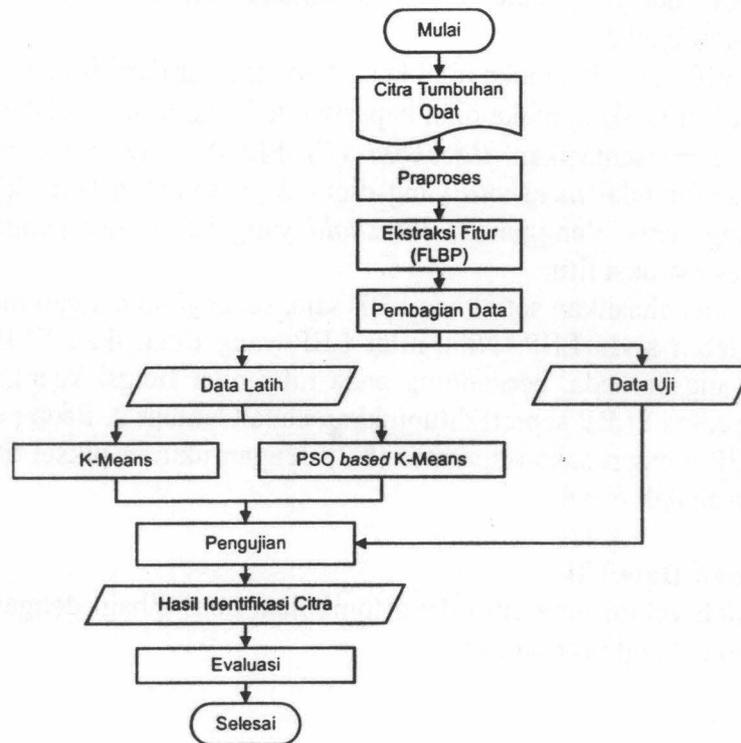
Penelitian ini bertujuan melakukan optimasi *k-means clustering* menggunakan metode PSO untuk meningkatkan hasil identifikasi tumbuhan obat.

## METODE

Penelitian ini menggunakan pendekatan secara sistem. Ekstraksi fitur pada citra menggunakan *fuzzy local binary pattern* (FLBP). Hasil ekstraksi berupa ciri dari citra akan dikelompokkan menggunakan *k-means clustering* dan PSO based *k-means clustering* untuk identifikasi citra. Metode penelitian yang akan dilakukan seperti ditunjukkan pada Gambar 1.

### Data Citra Tumbuhan Obat

Proses pengambilan citra daun tumbuhan obat dilakukan dengan pemotretan tiga puluh jenis tumbuhan obat di kebun Biofarmaka IPB dan di rumah kaca Pusat Konservasi Ex-Situ Tumbuhan Obat Hutan Tropika Indonesia, Fahutan, IPB. Pemotretan dilakukan dengan menggunakan lima kamera digital yang berbeda (DSC-W55, 7210 Supernova, Canon Digital Axus 95 IS, Samsung PL100, dan EX-Z35). Total citra daun tumbuhan obat yang digunakan adalah 1440 yang terdiri atas 30 spesies dengan masing-masing spesies terdiri atas 48 citra.



Gambar 1 Metode penelitian

### Praproses

Praproses adalah tahap awal dari proses identifikasi tumbuhan obat. Proses yang dilakukan pada tahap ini adalah mengganti latar belakang citra daun dengan warna putih. Ukuran citra diperkecil menjadi  $270 \times 240$  piksel. Mode warna citra diubah menjadi *grayscale* sebagai *input* untuk proses ekstraksi fitur.

### Ekstraksi Fitur dengan Fuzzy Local Binary Pattern

Menurut Iakovidis *et al.* (2008) proses *local binary pattern* (LBP) merepresentasikan tekstur lokal di sekitar tekstur pusat berdasarkan operator ketetanggaan LBP. Setiap pola tekstur LBP direpresentasikan oleh sembilan elemen  $P = \{P_{center}, P_0, P_1, \dots, P_7\}$ ,  $P_{center}$  merupakan nilai piksel pusat dan  $P_i$  ( $0 \leq i \leq 7$ ) merupakan nilai piksel sekelilingnya (*circular sampling*). Nilai *circular sampling* dapat dicirikan oleh nilai biner  $d_i$  ( $0 \leq i \leq 7$ ) yang kemudian akan dikonversi ke nilai desimal untuk mendapatkan nilai LBP dengan menggunakan Persamaan 1.

$$LBP = \sum_{i=0}^7 d_i \cdot 2^i, LBP \in [0, 255] \quad (1)$$

Nilai biner yang dihasilkan kemudian akan dikonversi ke nilai desimal untuk mendapatkan nilai LBP. Nilai-nilai LBP yang dihasilkan akan direpresentasikan melalui histogram. Menurut Ahonen *et al.* (2004) operator LBP dapat dikembangkan dengan menggunakan berbagai ukuran *sampling points* dan radius. Pengamatan piksel ketetanggaan, akan digunakan notasi  $(P, R)$  dimana  $P$  merupakan *sampling points* dan  $R$  merupakan radius. Nilai LBP dihasilkan sesuai dengan operator LBP yang digunakan. Semakin kecil radius dan semakin besar *sampling points* yang digunakan, semakin banyak piksel yang diolah untuk mendapatkan nilai LBP.

*Fuzzification* pada proses LBP adalah transformasi *variabel input* menjadi *variabel fuzzy* berdasarkan pada sekumpulan *fuzzy rule*. Berdasarkan penelitian Iakovidis *et al.* (2008),

penelitian ini menggunakan dua *fuzzy rule* untuk menentukan representasi nilai biner dan mencari nilai *fuzzy*. Kedua *rule* ialah:

*Rule R<sub>0</sub>*: semakin negatif nilai  $\Delta P_i$ , maka nilai kepastian terbesar dari  $d_i$  adalah 0.

*Rule R<sub>1</sub>*: semakin positif nilai  $\Delta P_i$ , maka nilai kepastian terbesar dari  $d_i$  adalah 1.

*Rule R<sub>0</sub>* dan *R<sub>1</sub>* merepresentasikan *threshold (F)* FLBP yang mengontrol derajat ketidakpastian. Semakin besar nilai *threshold* yang digunakan, semakin banyak nilai piksel yang diolah dalam rentang *fuzzy*. Penggunaan *threshold* yang besar mempengaruhi waktu komputasi pada saat proses ekstraksi fitur.

Metode LBP hanya menghasilkan satu kode LBP saja, sedangkan dengan metode FLBP menghasilkan satu atau lebih kode LBP. Nilai-nilai LBP yang dihasilkan FLBP memiliki tingkat kontribusi  $C_{LBP}$  yang berbeda, bergantung pada nilai-nilai fungsi keanggotaan yang dihasilkan. Ilustrasi dari proses FLBP seperti ditunjukkan pada Gambar 2. Pada penelitian ini, operasi LBP dengan FLBP menggunakan operator (8.2) dengan ukuran piksel 5x5 dan nilai *threshold* yang digunakan adalah  $F = 4$ .

**Pembagian Data Latih dan Data Uji**

Pada tahap ini jumlah vektor data citra daun tumbuhan obat dibagi dengan komposisi data latih sebesar 80% dan data uji sebesar 20%.

**K-Means Clustering**

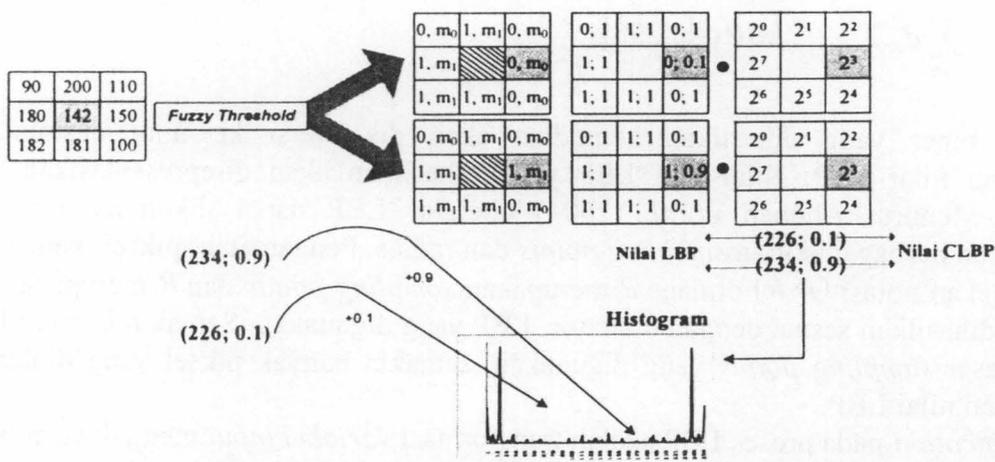
Algoritme *k-means clustering* adalah algoritme untuk mengklasifikasikan atau mengelompokkan data berdasarkan atribut-atribut atau fitur-fitur ke dalam beberapa jumlah kelompok ( $N_c$ ) secara tepat, dimana  $N_c$  adalah pusat *cluster*. Algoritme ini dimulai dengan menentukan pusat *cluster* awal secara acak. Setelah itu, objek data dalam *dataset* dikelompokkan ke pusat *cluster* berdasarkan jarak antara objek data dan pusat *cluster*. Algoritme *k-means clustering* (MacQueen 1967) terdiri dari pendefinisian pusat *cluster* awal, dan pengulangan dua langkah berikut ini:

(a) Pengelompokkan data ke dalam *cluster*

Pada langkah ini, data dimasukkan ke dalam salah satu *cluster* yang mempunyai pusat *cluster* terdekat dengan data tersebut. Dalam menghitung jarak antara data ke pusat *cluster* digunakan rumus Euclidean, didefinisikan pada Persamaan 2.

$$d(z_p, m_j) = \sqrt{\sum_{k=1}^{N_p} (z_{pk} - m_{jk})^2} \tag{2}$$

dengan  $N_p$  menunjukkan jumlah piksel citra,  $z_p$  menunjukkan komponen  $N_p$  piksel  $p$ ,  $m_j$  menunjukkan rata-rata pusat *cluster*  $j$ .



Gambar 2 Skema perhitungan FLBP dengan  $F = 10$

(b) Mencari pusat *cluster* yang baru

$$m_j = \frac{1}{n_j} \sum_{z_p \in C_{ij}} z_p \quad (3)$$

$n_j$  adalah jumlah data yang dimiliki *cluster j*. Keanggotaan data ( $C_{ij}$ ) yang masuk ke *cluster* akan dihitung untuk mendapatkan pusat *cluster* baru.

### Particle Swarm Optimization

Algoritme *particle swarm optimization* (PSO) diperkenalkan oleh Kennedy dan Eberhart (1995). Algoritme ini diinspirasi oleh perilaku sosial dari binatang seperti sekumpulan burung dalam suatu kumpulan (*swarm*). Perilaku sosial terdiri atas tindakan individu dan pengaruh dari individu-individu lain dalam satu kelompok. Setiap partikel di dalam PSO juga berhubungan dengan suatu kecepatan (*velocity*). Partikel-partikel mempunyai kecenderungan untuk bergerak ke area penelusuran yang lebih baik setelah melewati proses penelusuran.

Pada algoritme PSO, vektor kecepatan diperbaharui untuk masing-masing partikel, kemudian menjumlahkan vektor kecepatan tersebut ke posisi partikel. Proses memperbaharui kecepatan dipengaruhi oleh kedua solusi, yaitu melakukan penyesuaian posisi terbaik dari partikel (*particle best*) dan penyesuaian terhadap partikel terbaik dari seluruh kawan (*global best*). Pada tiap iterasi, setiap solusi yang direpresentasikan oleh posisi partikel dievaluasi dengan cara memasukkan solusi tersebut dalam *fitness function* (Santosa dan Willy 2011). Prosedur algoritme PSO ialah: (1) Inisialisasi populasi dari partikel-partikel dengan posisi dan kecepatan secara acak; (2) Evaluasi nilai *fitness* untuk masing-masing partikel; (3) Perbandingan dan pembaharuan *particle best* dan *global best* untuk tiap-tiap partikel berdasarkan fungsi *fitness*. Tahap selanjutnya adalah pengulangan langkah berikut sampai *stopping criteria* terpenuhi:

- a Menggunakan *particle best* dan *global best* yang ada, perbaharui kecepatan setiap partikel. Dari kecepatan yang diperoleh, perbaharui posisi partikel dengan menjumlahkan kecepatan baru dan posisi sebelumnya menggunakan Persamaan 4 dan Persamaan 5.

$$v_i(t) = wv_i(t-1) + c_1r_1(x_{pi} - x_i) + c_2r_2(x_{gi} - x_i) \quad (4)$$

$$x_i(t) = x_i(t-1) + v_i(t) \quad (5)$$

dengan

$i$  = indeks partikel

$t$  = iterasi

$w$  = *inertia*

$v_i$  = kecepatan partikel ke- $i$

$x_i$  = posisi partikel ke- $i$

$x_{gi}$  = posisi terbaik dari semua partikel (*gbest*)

$x_{pi}$  = posisi terbaik dari partikel ke- $i$  (*pbest*)

$c_{1,2}$  = *learning rate*

$r_{1,2}$  = bilangan acak [0,1]

- b Evaluasi *fitness* dari setiap partikel
- c Bandingkan dan perbaharui *particle best* dan *global best* tiap-tiap partikel berdasarkan fungsi *fitness*
- d Cek *stopping criteria*. Jika terpenuhi, berhenti. Jika tidak, kembali ke (a)

### PSO based *k*-means clustering

Algoritme PSO based *k*-means clustering merupakan gabungan dari algoritme *k*-means clustering dengan algoritme PSO. Pengelompokan data menggunakan algoritme PSO based *k*-means clustering menunjukkan hasil yang lebih baik dari metode *k*-means clustering (Van der Merwe dan Engelbrecht 2003). Di dalam konteks clustering, satu partikel merepresentasikan satu titik data pusat cluster. Setiap partikel  $x_i = (m_{i1}, \dots, m_{ij}, \dots, m_{iN_c})$  dengan  $m_{ij}$  mengacu pada  $j$ -th vektor pusat cluster dari partikel ke- $i$ . Input dalam algoritme PSO based *k*-means clustering adalah populasi data yang akan dikelompokkan, jumlah cluster ( $N_c$ ) dan jumlah iterasi yang diinginkan. Prosedur pertama dalam algoritme PSO based *k*-means clustering adalah inisialisasi yang meliputi inisialisasi jumlah cluster, populasi data, iterasi maksimum, serta pendefinisian pusat cluster awal. Kemudian, jarak antara objek data dan pusat cluster (Euclidean distance) dihitung, lalu data dikelompokkan ke dalam cluster. Untuk menghitung nilai fitness, kualitas dari setiap partikel diukur menggunakan Persamaan 6

$$f(x_i, Z) = \omega_1 \bar{d}_{\max}(Z, x_i) + \omega_2 (z_{\max} - d_{\min}(x_i)) \quad (6)$$

dengan  $z_{\max}$  adalah nilai maksimum piksel dalam set citra ( $z_{\max} = 2^s - 1$  untuk  $s$ -bit citra);  $Z$  adalah matriks yang mewakili pengolahan piksel citra ke cluster partikel  $I$  di mana setiap elemen  $z_{ijp}$  menunjukkan jika piksel  $z_p$  adalah milik cluster  $C_{ij}$  partikel  $I$ ; nilai  $\omega_1$  dan  $\omega_2$  ditentukan oleh user;

$$\bar{d}_{\max}(Z, x_i) = \max_{j=1, \dots, N_c} \left\{ \sum_{z_p \in C_j} d(z_p, m_j) / |C_{ij}| \right\} \quad (7)$$

adalah nilai rata-rata maksimum jarak Euclidean dari partikel untuk masing-masing kelas partikel dan

$$d_{\min}(x_i) = \min_{\forall j1, j2, j1 \neq j2} \{d(m_{ij1}, m_{ij2})\} \quad (8)$$

adalah jarak Euclidean minimum antara setiap pasang cluster.

Fungsi fitness tersebut memiliki tujuan untuk meminimalkan *intra-distance*, yaitu jarak antara data dengan masing-masing pusat cluster, seperti yang diukur dengan  $\bar{d}_{\max}(Z, x_i)$ , dan memaksimalkan *inter-distance*, yaitu jarak antara setiap pusat cluster, seperti yang diukur dengan  $d_{\min}(x_i)$ .

Dalam algoritme PSO based *k*-means clustering, yang digunakan adalah standar *gbest* PSO. Data dikelompokkan dengan dua langkah, yaitu (1) inisialisasi setiap partikel ( $k$ ) secara acak, (2) untuk  $t=1$  sampai  $t_{\max}$  lakukan langkah berikut:

- (a) Untuk setiap partikel  $i$  lakukan langkah berikut:
  - (b). Untuk setiap data vektor  $z_p$ 
    - (i). Hitung jarak Euclidean  $d(z_p, m_j)$  untuk semua pusat cluster  $C_{ij}$
    - (ii). Tentukan  $(x, y)$  ke cluster  $C_{ij}$  sehingga  $d(x, y) = \text{minimum}$
    - (iii). Hitung nilai fitness menggunakan persamaan (6)
  - (b) Perbaharui *pbest* dan *gbest*
  - (c) Perbaharui pusat cluster menggunakan persamaan (4) dan (5).

Kembali ke langkah 2 sampai kriteria terpenuhi, proses berhenti ketika tidak ada data yang berpindah cluster atau sampai jumlah maksimum iterasi yang ditentukan.

### Pengujian

Pengujian untuk model *k*-means clustering dan PSO based *k*-means clustering dilakukan dengan menggunakan metode pengukuran jarak Euclidean, menggunakan persamaan (2).

### Evaluasi Model

Evaluasi dari kinerja model *clustering* terdiri atas 2 bagian yaitu evaluasi nilai *quantization error*, didefinisikan pada Persamaan 9 dan evaluasi pada banyaknya data uji yang diprediksi secara benar atau salah oleh model, dihitung menggunakan akurasi yang didefinisikan pada Persamaan 10.

$$J_e = \frac{\sum_{j=1}^{N_c} \left[ \sum_{\forall z_p \in C_{ij}} d(z_p, m_j) / |C_{ij}| \right]}{N_c} \quad (9)$$

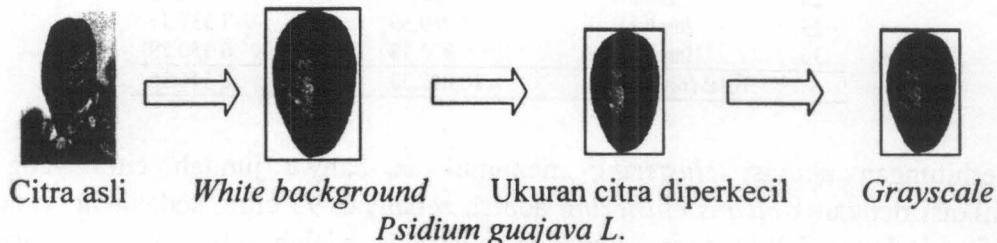
dengan  $d(z_p, m_j)$  adalah jumlah vektor data yang masuk pada *cluster*  $C_{ij}$  dan  $N_c$  adalah jumlah cluster yang ditentukan.

$$\text{akurasi} = \frac{\text{banyaknya prediksi data uji yang benar}}{\text{total data uji}} \times 100\% \quad (10)$$

## HASIL DAN PEMBAHASAN

### Praproses

Praproses data bertujuan untuk mengurangi waktu pemrosesan data (*running time*), seperti ditunjukkan pada Gambar 3.



Gambar 3 Hasil praproses citra tumbuhan obat

### Ekstraksi Fitur FLBP

Citra hasil praproses diekstrak menggunakan nilai *threshold* FLBP fuzzifikasi ( $F$ ) yaitu 4 dengan operator LBP (8.2). Panjang *bin* yang dihasilkan pada histogram  $FLBP_{P,R}$  bergantung pada jumlah *sampling points* ( $P$ ) yang digunakan, yaitu  $2^P$ . Jumlah  $P$  yang digunakan adalah 8 sehingga jumlah *bin* pada histogram  $FLBP_{P,R}$  sebanyak  $2^8 = 256$  bin.

### Hasil Percobaan

Pada tahap ini, metode *k-means clustering* tanpa optimasi dibandingkan dengan yang menggunakan metode optimasi PSO. Metode PSO based *k-means clustering* menggunakan parameter *learning rate*  $c_1 = 1.49$  dan  $c_2 = 1.49$ , nilai  $r_1 = 0.5$  dan  $r_2 = 0.5$ , untuk bobot inerti ( $w$ ) = 0.72 dengan kecepatan maksimum ( $V_{max}$ ) = 2. Kualitas *clustering* untuk kedua metode diukur dengan nilai *quantization error* seperti ditunjukkan pada Tabel 1. Nilai rata-rata *quantization error* untuk *k-means clustering* adalah sebesar 10072.33, sedangkan untuk PSO based *k-means clustering* sebesar 8146.35. Pada percobaan tersebut, terlihat bahwa metode PSO based *k-means clustering* lebih baik dalam proses pengelompokan data citra daun tumbuhan obat karena mempunyai nilai *quantization error* yang lebih kecil.

Tabel 1 Perbandingan nilai *quantization error*

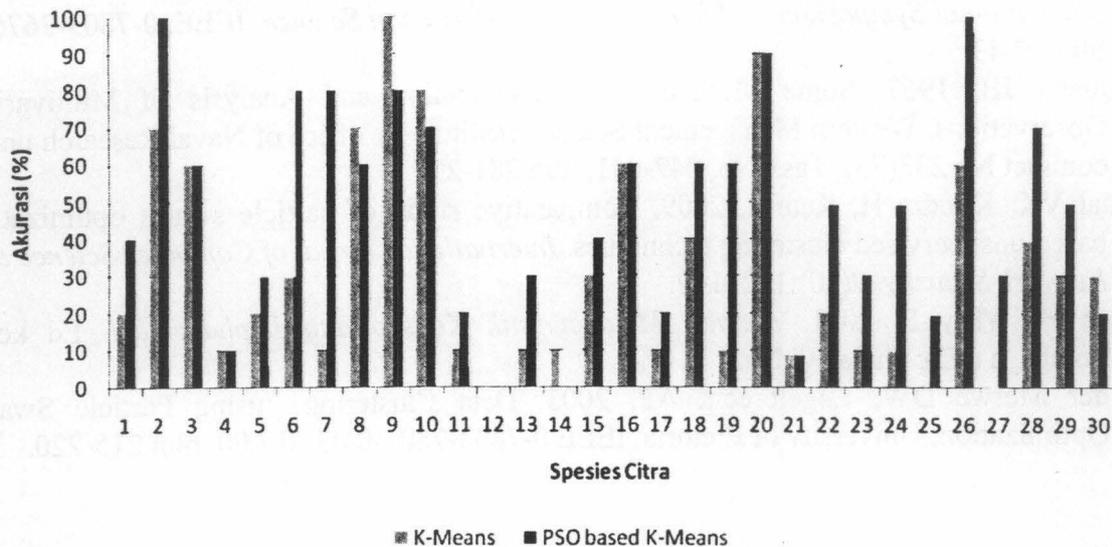
No	Citra	K-means	PSO based k-means
1	Pandan Wangi	4 339.65	4 934.39
2	Jarak Pagar	19 661.80	13 904.40
3	Dandang Gendis	4 364.76	5 156.08
4	Lavender	3 806.62	5 057.63
5	Akar Kuning	14 374.70	10 722.80
6	Daruju	10 170.90	8 143.24
7	Pegagan	8 871.66	8 776.05
8	Andong	5 513.26	5 174.42
9	Kemangi	9 059.36	5 804.22
10	Iler	9 855.74	8 718.72
11	Jeruk Nipis	8 866.15	8 258.82
12	Bidani	9 634.54	7 193.98
13	Gadung Cina	11 800.00	8 913.41
14	Tabat Barito	12 499.10	9 663.39
15	Nandang gendis kuning	8 135.32	6 719.57
16	Bunga Telang	10 304.60	7 884.19
17	Mangkokan	16 577.40	12 204.10
18	Som Jawa	6 462.30	6 337.01
19	Pungpulutan	14 130.30	9 997.45
20	Sosor Bebek	5 871.55	5 672.69
21	Nanas kerang	6 096.82	5 725.27
22	Seligi	11 042.10	9 305.69
23	Remak Daging	18 189.00	13 418.00
24	Kumis Kucing	10 905.30	8 144.93
25	Kemuning	12 039.00	7 394.52
26	Cincau Hitam	23 126.20	16 019.40
27	Sambang Darah	6 431.79	5 793.87
28	Landik	4 226.14	5 464.62
29	Jambu Biji	8 910.59	7 537.37
30	Handeuleum	6 903.18	6 350.29
Total (rata-rata)		10 072.33	8 146.35

Perhitungan akurasi *clustering* menunjukkan bahwa jumlah citra yang berhasil diidentifikasi dengan *k-means clustering* adalah sebanyak 95 citra, sedangkan yang berhasil diidentifikasi dengan *PSO based k-means clustering* adalah sebanyak 142 citra, dengan akurasi sebagai berikut:

$$Kmeansclustering = \frac{95}{300} \times 100\% = 31.67\%$$

$$PSObasedkmeansclustering = \frac{142}{300} \times 100\% = 47.33\%$$

Perbandingan akurasi hasil *clustering* untuk masing-masing spesies citra ditunjukkan pada Gambar 4. Metode *k-means clustering* menunjukkan bahwa terdapat 27 spesies citra yang dapat diidentifikasi dan 3 spesies yang tidak dapat diidentifikasi. Spesies 9 (Kemangi) memiliki tingkat akurasi 100%, sedangkan spesies 12 (Bidani), 25 (Kemuning) dan 27 (Sambang Darah) tidak teridentifikasi sehingga nilai akurasinya 0%. Pada metode *PSO based k-means clustering*, terdapat 28 spesies yang teridentifikasi dengan baik dan 2 spesies yang tidak teridentifikasi. Spesies 2 (Jarak Pagar) dan Spesies 26 (Cincau Hitam) merupakan spesies yang berhasil diidentifikasi dengan tingkat akurasi 100%, sedangkan spesies 12 (Bidani) dan spesies 14 (Tabat Barito) tidak dapat teridentifikasi dan menghasilkan akurasi 0%.



Gambar 4 Grafik perbandingan akurasi per spesies citra

Dari nilai akurasi yang diperoleh, yaitu 31.67% untuk *k-means clustering* dan 47.33% untuk *PSO based k-means clustering*, dapat disimpulkan bahwa metode *PSO based k-means clustering* lebih baik daripada metode *k-means clustering*. Kesalahan identifikasi disebabkan oleh tingkat kemiripan citra daun tumbuhan obat yang tinggi karena tekstur daun yang hampir sama. Kemiripan ini membuat akurasi identifikasi menjadi rendah, karena banyak citra yang teridentifikasi sebagai spesies yang berbeda.

## SIMPULAN

Optimasi *k-means clustering* menggunakan metode PSO pada citra tumbuhan obat berhasil diimplementasikan. Metode ini menggunakan data masukan hasil ekstraksi fitur FLBP berupa vektor ciri dari citra daun tumbuhan obat.

Metode *PSO based k-means clustering* lebih baik daripada metode *k-means clustering* karena memiliki nilai *quantization error* yang lebih kecil. Nilai *quantization error* untuk metode *PSO based k-means clustering* adalah sebesar 8146.35, sedangkan untuk metode *k-means clustering* adalah sebesar 10 072.33. Kesimpulan ini didukung dengan hasil akurasi yang diperoleh, yaitu 47.33% untuk *PSO based k-means clustering* dan 31.67% untuk *k-means clustering*.

## DAFTAR PUSTAKA

- Ahonen T, Hadid A, Pietikainen M. 2004. Face recognition with local binary pattern. Springer Verlag Berlin Heidelberg, ECCV, LNCS 3021, hlm 469-481.
- Groombridge B, Jenkins MD. 2002. *World Atlas of Biodiversity*. Earth's Living Resources in the 21<sup>st</sup> Century, UNEP-WCMC, Cambridge, 0-520-23668-8.
- Herdiyeni Y, Wahyuni NKS. 2012. Mobile application for Indonesian medicinal plants identification using fuzzy local binary pattern and fuzzy color histogram. Di dalam: *International Conference on Advanced Computer Science and Information System*. hlm 1-31.
- Iakovidis DK, Eystratios G, Keramidas EG, Maroulis D. 2008. Fuzzy local binary patterns for ultrasound texture characterization, ICIAR, LNCS 5112, pp. 750-759.
- Jain AK, Murty MN, Flynn PJ. 1999. Data clustering: a review. *ACM Computing Surveys*. 31(3):265-323.

- Kennedy J, Eberhart R. 1995. A new optimizer using particle swarm theory. Di dalam: *Sixth International Symposium on Micro Machine and Human Science*. IEEE. 0-7803-2676-8. hlm 39-43.
- MacQueen JB. 1967. Some Methods for Classification and Analysis of Multivariate Observations. Western Management Science Institute by office of Naval Research under contract No 233(75), Task No. 047-041, hlm 281-297.
- Panchal VK, Kundra H, Kaur J. 2009. Comparative study of particle swarm optimization based unsupervised clustering techniques. *International Jurnal of Computer Science and Network Security*. 9(10):132-140.
- Santosa B, Willy P. 2011. *Metoda Metaheuristik Konsep dan Implementasi*. Ed ke-1. Surabaya (ID): Guna Widya.
- Van der Merwe DW, Engelbrecht AP. 2003. Data Clustering using Particle Swarm Optimization, University of Pretoria, IEEE 0-7803-7804-0/03/\$17.00, hlm 215-220.