

Metode SVM untuk Klasifikasi Enam Tumbuhan *Zingiberaceae* Menggunakan Variabel Terpilih Hasil Algoritma Genetika *

Triyani Oktaria¹, Utami Dyah Syafitri^{2‡}, Mohamad Rafi³, Farit M Afendi⁴

¹Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, Indonesia

²³⁴Department of Statistics, IPB University, Indonesia

[‡]corresponding author: utamids@apps.ipb.ac.id

Copyright © 2021 Triyani Oktaria, Utami Dyah Syafitri, Mohamad Rafi, and Farit M Afendi. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Ginger, red ginger, emprit ginger, elephant ginger, red galangal and white galangal are known to have similar shapes and uses, especially those that are packaged in powder form. In this study, UV-Vis spectrum 200nm-700nm were used as a source of data from chemical compound contain in those plants for classification of the six plants. In this research, the support vector machine (SVM) classification method was used to classify the six plants. Another goal of this study was to identify the wavelengths which give more information about the chemical compound of the plants. The preprocessing procedure was implemented by construction of a genetic algorithm. There were four parameters in the genetic algorithm were set namely population size, crossover probability, mutation, and generation probability. The mutation and the population size influenced significantly the results of SVM. The best result was given by probability of mutation was 10 and population size was 30. The SVM model was better than the SVM model without preprocessing procedure.

Keywords: UV-Vis, genetic algorithm, support vector machine, preprocessing.

1. Pendahuluan

Beberapa jenis tanaman rimpang seperti jahe emprit (*Zingiber officinale* var *amarum*), jahe merah (*Z. officinale* var *rubrum*), lempuyang emprit (*Z. americans*), lempuyang gajah (*Z. zerumbet*), lengkuas merah (*Alpinia purpurata*), dan lengkuas putih (*A. galanga*) memiliki bentuk dan kegunaan yang mirip. Adanya kemiripan bentuk dan kegunaan yang ada pada tanaman rimpang ini dimungkinkan terjadinya substitusi bahan baku antara satu dengan lainnya sehingga sulit dibedakan, misalnya dalam produk obat herbal, terutama jika dikemas dalam bentuk serbuk. Pusat Studi

* Received: May 2021; Reviewed: May 2021; Published: May 2021

Biofarmaka Tropika Institut Pertanian Bogor melakukan percobaan terhadap enam tanaman rimpang tersebut dengan menggunakan spektrofometri UV-Vis (Ultraviolet-Visible) untuk mengidentifikasi kandungan senyawa aktif yang ada pada tanaman tersebut. Hafid (2020) telah melakukan penelitian pengklasifikasian terhadap jahe emprit, jahe merah, lempuyang emprit, lempuyang gajah, lengkuas merah, dan lengkuas putih menggunakan analisis diskriminan kuadrat terkecil parsial tanpa preprocessing diperoleh nilai akurasi 86%, sedangkan dengan menggunakan analisis diskriminan kuadrat terkecil parsial dengan preprocessing diperoleh nilai akurasi 95%.

Seiring dengan perkembangan teknologi yang semakin cepat dan pesat, semakin banyak metode klasifikasi yang dikembangkan dalam mengumpulkan, mengelola, dan menyimpan data yang berukuran besar serta stabil terhadap pencilan. Pada prinsipnya, klasifikasi merupakan proses mengelompokkan data ke dalam suatu kelas. Data yang berada pada kelas yang sama relatif memiliki karakteristik atau informasi yang sama. Salah satu metode klasifikasi yang dapat digunakan adalah support vector machine (SVM) untuk memprediksi kelas dari suatu data. Metode SVM telah diterapkan dalam berbagai bidang, termasuk dalam bidang kemometrik. Chen et al. (2007) menggunakan metode SVM dalam mengklasifikasikan data spektrum keluaran spektroskopi, hasil yang diperoleh menunjukkan bahwa rata-rata akurasi pendugaan mencapai 95%. Metode SVM juga memiliki kemampuan klasifikasi yang baik dan dapat digunakan untuk data berskala besar dengan peubah penjelas yang banyak serta kekar terhadap pencilan (Steinberg & Colla 1995).

Kebaikan hasil prediksi klasifikasi dipengaruhi oleh peubah penjelas, sehingga pemilihan peubah penjelas berperan penting dalam memperoleh hasil prediksi dengan akurasi yang tinggi, terlebih sangat banyak peubah penjelas yang digunakan. Salah satu teknik identifikasi pemilihan peubah penjelas yang optimum dapat menggunakan teknik algoritma genetika. Teknik ini juga sering digunakan dalam bidang kemometrika. Leardi (2001) memberikan ulasan tentang penggunaan algoritma genetika di bidang kimia khususnya dalam kemometrika, Sartono (2010) berhasil menerapkan algoritma genetika dalam pemilihan peubah penjelas pada analisis regresi yang terdiri atas ratusan peubah. Selain itu, Khempila dan Boonjing (2011) menggunakan algoritma genetika untuk mereduksi jumlah peubah penyakit jantung koroner.

Penelitian ini menggunakan metode SVM untuk mengklasifikasikan enam tanaman rimpang yang terdiri dari jahe emprit, jahe merah, lempuyang emprit, lempuyang gajah, lengkuas merah dan lengkuas putih dengan identifikasi pemilihan peubah penjelas panjang gelombang optimum dengan menggunakan algoritma genetika.

2. Metodologi

Metode yang digunakan dalam penelitian ini adalah metode Support Vector Machine (SVM) dengan identifikasi pemilihan peubah penjelas panjang gelombang optimum dengan menggunakan algoritma genetika.

2.1 Bahan dan Data

Data yang digunakan dalam penelitian ini mengacu pada penelitian yang telah dilakukan oleh Hafid (2019) yaitu data primer hasil percobaan yang dilakukan oleh Pusat Studi Biofarmaka Tropika Institut Pertanian Bogor pada tahun 2014 terhadap produk herbal serbuk yang terdiri dari jahe emprit (JE), jahe merah (JM), lempuyang emprit (LE), lempuyang gajah (LG), lengkuas merah (LM) dan lengkuas putih (LP). Total sampel sebanyak 124 yang terdiri dari 23 sampel jahe emprit, 10 sampel jahe merah, 22 sampel lempuyang emprit, 11 sampel lempuyang gajah, 29 sampel lengkuas merah, dan 29 sampel lengkuas putih. Respon percobaan merupakan nilai absorban yang diuji menggunakan spektrofometer ultraviolet visible (UV-VIS) pada panjang gelombang 240 nm-700 nm dengan total 1.001 titik panjang gelombang.

2.2 Metode Penelitian

Tahapan analisis pada penelitian ini sebagai berikut:

A. Eksplorasi Data

1. Mendeteksi pencilan dengan membuat plot garis antara nilai absorban dan panjang gelombang.
2. Penanganan data pencilan. Penanganan data pencilan dan titik potong panjang gelombang didasarkan pada penelitian yang dilakukan oleh Hafid H (2020).

B. Membagi data menjadi dua yaitu data training dan data testing dengan persentase 70%: 30%

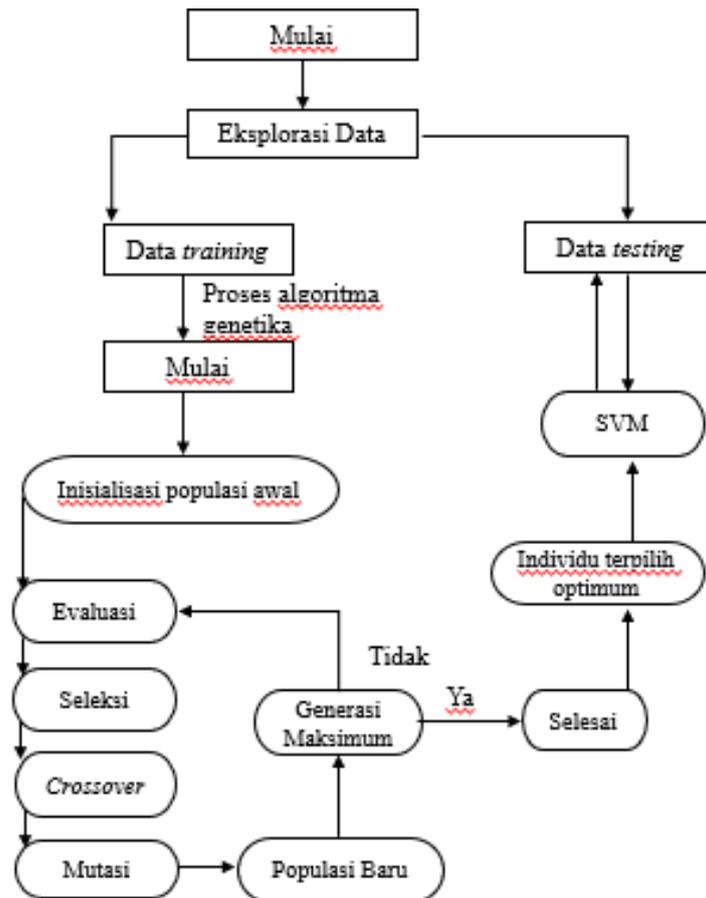
C. Mencari parameter terbaik pada SVM dengan melakukan optimasi pada parameter C. Parameter C dipilih untuk mengontrol nilai kesalahan pada klasifikasi. Parameter C ditentukan dengan mencoba beberapa nilai dan dievaluasi galatnya. Nilai parameter C yang diujicobakan adalah $1e-05$ s.d. $1e+05$

D. Melakukan rancangan simulasi dengan mengkombinasikan parameter-parameter algoritma genetika seperti, ukuran populasi, peluang crossover, peluang mutasi, dan generasi dengan rincian pada Tabel 1.

Tabel 1: Rancangan simulasi parameter algoritma genetika

No	Parameter	Nilai Parameter
1	Ukuran Populasi	10, 30
2	Crossover	0,1; 0,3; 0,5
3	Peluang Mutasi	0,1; 0,4; 0,7
4	Generasi	5, 10
5	Iterasi	10

E. Melakukan proses komputasi algoritma genetika dengan evaluasi performansi peubah terpilih dengan menggunakan nilai akurasi. Adapun prosedur komputasi algoritma genetika dijelaskan pada Gambar 1.



Gambar 1: Prosedur komputasi algoritma genetika.

- F. Melakukan analisis ragam pengaruh faktor utama rancangan faktorial terhadap nilai akurasi.
- G. Identifikasi faktor utama yang berpengaruh terhadap model yang dibangun.

3. Hasil dan Pembahasan

3.1 Eksplorasi Data

Eksplorasi data spektrum UV-Vis jahe emprit, jahe merah, lempuyang emprit, lempuyang gajah, lengkuas merah dan lengkuas putih dilakukan dengan membuat plot. Plot data spektrum awal digunakan untuk mengetahui pola data. Pola data enam tanaman rimpang menunjukkan ada beberapa amatan spektrum yang diduga pencilan sehingga perlu dilakukan penanganan terhadap data pencilan tersebut. Penanganan pencilan dilakukan dengan menghilangkan amatan pencilan pada panjang gelombang tertentu. Hafid (2019) melakukan penanganan pencilan dengan melakukan pemotongan panjang gelombang pada 240-300nm sehingga diperoleh sampel sebanyak 113 sampel terdiri atas 20 sampel jahe emprit, 10 sampel jahe merah, 18 lempuyang emprit, 10 lempuyang gajah, 27 lengkuas merah dan 28 lengkuas putih.

3.2 Metode Support Vector Machine (SVM)

3.2.1. Pembentukan Model Awal SVM

Masalah dasar dari SVM adalah menentukan nilai parameter yang mampu menghasilkan bidang pembatas yang memisahkan data antarkelas dengan baik. Nilai parameter yang tidak tepat dapat menghasilkan akurasi yang rendah. Pembentukan model awal pada SVM dilakukan dengan cara tuning parameter pada nilai C 1e-05 s.d. 1e+05 yang bertujuan untuk mencari parameter yang terbaik. Parameter terbaik dilihat dari nilai galat yang terkecil. Berdasarkan Tabel 2, diperoleh bahwa nilai parameter terbaik berada pada C=10.000 dengan nilai galat 0.2625. Selanjutnya, pemodelan dilakukan dengan nilai C=10.000 untuk seluruh kombinasi pemodelan yang dibangun.

Tabel 2: Tuning parameter SVM

No	C	Galat
1	1e-05	0.8250
2	1e-04	0.8250
3	1e-03	0.6250
4	1e-02	0.5000
5	1e-01	0.3875
6	1e+00	0.4125
7	1e+01	0.3000
8	1e+02	0.3125
9	1e+03	0.2750
10	1e+04	0.2625
11	1e+05	0.2625

3.2.2. Klasifikasi Data Menggunakan SVM

Pembagian data sampel dilakukan terlebih dahulu untuk klasifikasi sampel menggunakan model SVM dengan terlebih dahulu melakukan pembagian data. Data sampel dibagi menjadi dua bagian yaitu data latih sebanyak 70% dan data sebesar 30%. Berdasarkan data latih diterapkan metode SVM linear dan SVM radial tanpa preprocessing dengan nilai parameter C=10.000 untuk melihat metode SVM yang tepat untuk mengklasifikasikan keenam tanaman rimpang. Hasil klasifikasi menggunakan metode SVM linear tanpa preprocessing diperoleh nilai akurasi yang lebih tinggi sebesar 90,91% dibandingkan dengan metode SVM radial tanpa preprocessing dengan akurasi sebesar 78,79%. Klasifikasi prediksi sampel menggunakan data uji sampel menggunakan metode SVM linear tanpa preprocessing ditunjukkan pada Tabel 3. Data uji sampel yang digunakan sebanyak 33 sampel dengan rincian 3 JM, 6 JE, 6 LE, 2 LG, 8 LM, dan 8 LP. Berdasarkan hasil prediksi menunjukkan bahwa terdapat misclassification untuk jahe merah, lempuyang empريت, dan lengkuas merah.

Tabel 3: Matriks klasifikasi SVM linear tanpa preprocessing

Hasil Prediksi	Data Aktual					
	JM	JE	LE	LG	LM	LP
JM	2	0	0	0	0	0
JE	1	6	0	0	0	0
LE	0	0	5	0	0	0
LG	0	0	1	2	0	0
LM	0	0	0	0	7	0
LP	0	0	0	0	1	8

Nilai akurasi SVM linear tanpa preprocessing yang sebesar 90,91% sudah cukup tinggi, namun kurang efektif karena tanpa menggunakan seleksi peubah penjelas. Seleksi peubah penjelas biasa digunakan ketika jumlah peubah penjelas yang digunakan sangat banyak dan ada indikasi korelasi yang tinggi. Model yang baik ditunjukkan dengan nilai akurasi yang tinggi dan peubah penjelas yang sedikit sehingga mudah untuk diinterpretasikan. Berdasarkan hasil diatas, maka dilakukan seleksi peubah penjelas dengan menggunakan metode algoritma genetika untuk mengidentifikasi peubah penjelas absorban di panjang gelombang mana yang memberikan informasi yang optimum.

3.2.3. SVM-Algoritma Genetika

Metode algoritma genetika merupakan salah satu metode optimasi yang dibangun mengikuti pola evolusi pada makhluk hidup guna memperoleh suatu solusi yang terbaik. Pada penelitian ini, algoritma genetika digunakan untuk mengidentifikasi panjang gelombang terpilih. Selanjutnya, absorban pada panjang gelombang terpilih digunakan dalam membangun SVM untuk klasifikasi enam tanaman rimpang yang digunakan yaitu jahe emprit, jahe merah, lempuyang emprit, lempuyang gajah, lengkuas merah dan lengkuas putih. Proses pemilihan peubah penjelas panjang gelombang ini membutuhkan waktu yang banyak karena jumlah peubah penjelas yang digunakan sangat banyak. Pemilihan panjang gelombang yang optimum dalam algoritma genetika dilakukan dengan memberikan variasi pada parameter algoritma genetika seperti ukuran populasi, peluang crossover, peluang mutasi, dan generasi. Evaluasi proses algoritma genetika dilakukan dengan nilai akurasi.

Strategi algoritma genetika pada penelitian ini, beberapa parameter ditentukan diawal dan diacu dalam proses komputasi algoritma genetika sebagaimana pada Tabel 1. Parameter ukuran populasi ditentukan sebesar 10 dan 30, peluang crossover ditentukan sebesar 0,1, 0,3, dan 0,5. Peluang mutasi ditentukan sebesar 0,1, 0,4, dan 0,7, sedang parameter generasi ditentukan 5 dan 10. Iterasi pada variasi parameter dilakukan sebanyak 10 kali untuk memastikan bahwa data terpilih merupakan data yang acak. Sehingga total pemodelan yang dibangun sebanyak 360 model.

3.2.4. Analisis Ragam Pengaruh Faktor Utama Rancangan Faktorial

Analisis ragam rancangan faktorial bertujuan untuk mengetahui pengaruh faktor utama dan pengaruh interaksi. Faktor utama pada penelitian ini merupakan parameter-parameter algoritma genetika yang terdiri dari ukuran populasi, peluang crossover, peluang mutasi dan generasi. Hasil eksplorasi data, tidak terlihat adanya interaksi antar parameter, sehingga analisis ragam dilakukan fokus untuk melihat pengaruh faktor utama. Berdasarkan Tabel 4, terlihat bahwa pengaruh faktor peluang mutasi dan ukuran populasi nyata pada taraf 5%. Sedangkan pengaruh faktor utama peluang crossover dan generasi tidak nyata pada taraf 5%. Hal ini menunjukkan bahwa penentuan ukuran populasi diawal untuk inialisasi sangat penting dan berpengaruh pada perolehan peubah optimum. Hasil analisis ragam yang disajikan Tabel 4 belum mampu melihat kombinasi parameter mana yang paling baik terhadap nilai akurasi.

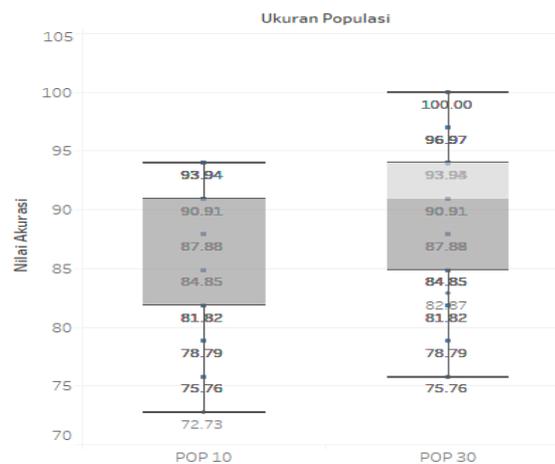
Tabel 4: Analisis ragam rancangan faktorial

Sumber Keragaman	Derajat Bebas	Jumlah Kuadrat (Adj)	Kuadrat Tengah (Adj)	Nilai-F	Nilai P
Peluang <i>crossover</i>	2	2,3	1,14	0,05	0,952
Peluang mutasi	2	3371,1	1685,56	72,12	0,000
Ukuran Populasi	1	647,8	647,80	27,72	0,000
Generasi	1	0,7	0,74	0,03	0,859
Galat	353	8250,3	23,37		
Total	359	12272,2			

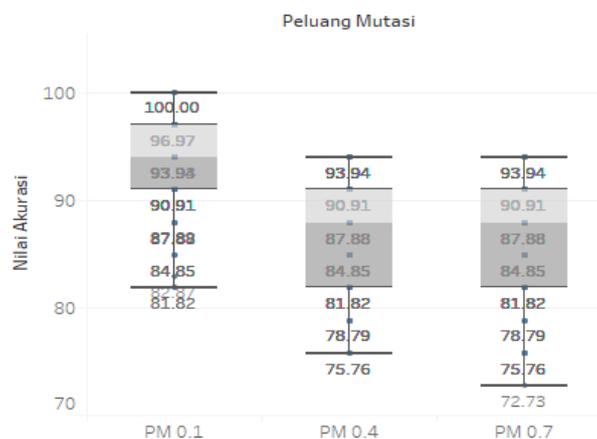
3.2.5. Pengaruh Parameter Algoritma

Berdasarkan Tabel 4, menunjukkan bahwa parameter ukuran populasi dan peluang mutasi yang memberikan pengaruh terhadap nilai rata-rata akurasi. Hal ini juga menunjukkan bahwa rancangan ukuran populasi dan peluang mutasi yang ditentukan sangatlah berperan penting dalam proses pencarian solusi nilai optimal. Nilai optimal inilah yang menentukan peubah penjelas panjang gelombang terpilih yang selanjutnya digunakan untuk membangun model SVM. Algoritma genetika mulai bekerja dengan populasi awal yang ditentukan sebesar 10 dan 30. Eksplorasi data akurasi digambarkan dengan menggunakan boxplot. Salah satu kelebihan dari boxplot adalah dapat melihat keragaman sebaran data. Grafik boxplot yang ditunjukkan pada Gambar 2, menunjukkan bahwa gambaran rata-rata akurasi parameter ukuran populasi. Boxplot pada ukuran populasi 30 memiliki nilai rata-rata akurasi maksimum 100%, nilai rata-rata akurasi minimum 75,76%, rata-rata akurasi pada kuartil 1 (Q1) 84,85%, rata-rata akurasi pada kuartil 2 (Q2) 90,91%, dan rata-rata akurasi pada kuartil 3 (Q3) 93,94%. Sedangkan rentang rata-rata akurasi berada pada 72,73% – 100%.

Secara visual, rata-rata akurasi pada boxplot ukuran populasi 30 memiliki nilai rata-rata akurasi yang lebih tinggi dibandingkan ukuran populasi 10. Selain itu, pada proses komputasi algoritma genetika, ukuran populasi 30 cenderung lebih lama dibandingkan ukuran populasi 10 dengan generasi yang semakin meningkat. Proses komputasi diulang sebanyak 10 kali untuk memastikan bahwa individu yang terpilih merupakan individu yang acak. Setiap pengulangan hampir pada semua parameter mengalami penurunan nilai akurasi walaupun relatif kecil. Hal ini dapat terjadi karena setiap kali menjalankan program genetika tidak akan selalu menjamin hasil yang diperoleh akan sama dengan sebelumnya, hasilnya dapat lebih baik atau sebaliknya.



Gambar 2: Boxplot rata-rata akurasi parameter ukuran populasi



Gambar 3: Boxplot rata-rata akurasi parameter peluang mutasi

Selain parameter ukuran populasi, parameter yang berpengaruh adalah peluang mutasi. Peluang mutasi yang digunakan pada proses algoritma genetika adalah 0,1, 0,4, dan 0,7. Proses mutasi bertujuan untuk mendapatkan individu terbaik dengan cara memodifikasi satu atau lebih nilai gen dalam satu individu. Peluang mutasi 0,1 artinya proses modifikasi gen yang dilakukan sebesar 10%

dari total individu terpilih setelah proses pindah silang (crossover). Berdasarkan Gambar 3, boxplot rata-rata akurasi peluang mutasi menunjukkan bahwa rentang rata-rata akurasi untuk semua kombinasi parameter berada pada 72,73% – 100%. Peluang mutasi 0,1 memiliki keragaman rata-rata akurasi yang paling kecil dibandingkan peluang mutasi 0,4 dan 0,7. Hal ini menunjukkan bahwa 10% individu baru hasil mutasi mampu mengklasifikasikan keenam jenis tanaman rimpang jahe emprit, jahe merah, lempuyang emprit, lempuyang gajah, lengkuas merah dan lengkuas putih dengan baik. Nilai rata-rata akurasi minimum pada peluang mutasi 0,1 sebesar 81,02% dan nilai rata-rata akurasi maksimum sebesar 100%. Berdasarkan Gambar 3, menunjukkan pula bahwa semakin tinggi peluang mutasi, semakin besar keragaman akurasi yang diperoleh. Selain itu, nilai akurasi menunjukkan penurunan dengan meningkatnya peluang mutasi. Peluang mutasi 0,4 dan peluang mutasi 0,7 memiliki rata-rata akurasi maksimum yang sama sebesar 93,94%.

Metode SVM linear tanpa preprocessing telah menunjukkan nilai akurasi sebesar 90,91% dengan peubah penjelas panjang gelombang sebanyak 121. Hasil yang lebih baik telah ditunjukkan dengan melakukan preprocessing pemilihan panjang gelombang dengan menggunakan algoritma genetika. Berdasarkan Tabel 5 menunjukkan bahwa metode SVM dengan identifikasi panjang gelombang menggunakan algoritma genetika mampu mengklasifikasikan keenam tanaman rimpang jahe emprit, jahe merah, lempuyang emprit, lempuyang gajah, lengkuas merah dan lengkuas putih secara tepat dengan nilai akurasi 100%. Hal ini memberikan makna bahwa data prediksi sama dengan data yang sebenarnya, terklasifikasi sesuai dengan kelasnya seperti pada Tabel 5. Tiga jahe merah (JM) pada data aktual terprediksi sejumlah tiga jahe merah (JM) pada data hasil prediksi, delapan lengkuas putih (LP) pada data aktual terprediksi sejumlah delapan lengkuas putih (LP) pada data hasil prediksi. Interpretasi yang sama untuk jahe emprit (JE), lempuyang emprit (LE), lempuyang gajah (LG), dan lengkuas merah (LM).

Tabel 5: Confusion matrix klasifikasi enam tanaman rimpang

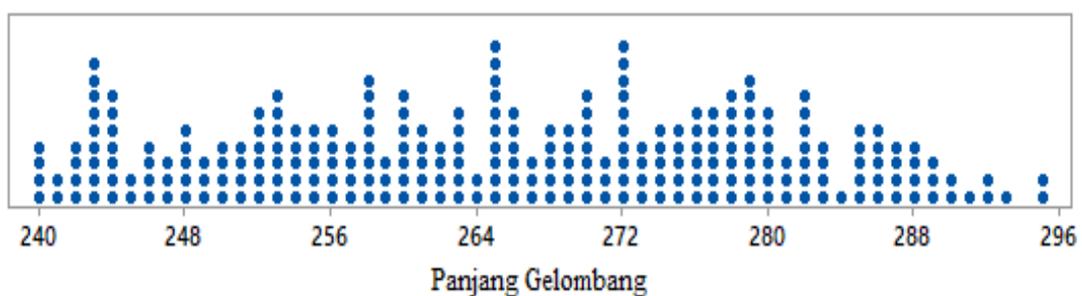
Hasil Prediksi	Data Aktual					
	JM	JE	LE	LG	LM	LP
JM	3	0	0	0	0	0
JE	0	6	0	0	0	0
LE	0	0	5	0	0	0
LG	0	0	0	3	0	0
LM	0	0	0	0	8	0
LP	0	0	0	0	0	8

Terdapat enam model dengan tingkat akurasi 100%, yaitu pada parameter peluang crossover, peluang mutasi, ukuran populasi, dan generasi yang dirinci

pada Tabel 6. Rentang panjang gelombang berada pada 240 nm-295,5 nm. Adapun jumlah peubah penjelas panjang gelombang yang terpilih pada tingkat akurasi 100% sebanyak 40 peubah penjelas pada parameter peluang crossover 0,1, peluang mutasi 0,1, ukuran populasi 30 dan generasi 10, sebanyak 35 peubah penjelas pada kombinasi parameter peluang crossover 0,1, peluang mutasi 0,1, ukuran populasi 30 dan generasi 10. Sebanyak 58 peubah penjelas pada kombinasi parameter peluang crossover 0,3, peluang mutasi 0,1, ukuran populasi 30 dan generasi 10. Jika dibandingkan dengan 360 model yang dibangun, maka metode algoritma genetika mampu mereduksi peubah penjelas sekitar 40% - 77% dari total 121 peubah penjelas. Jumlah minimum peubah penjelas terpilih sebanyak 28, yaitu pada peluang mutasi 0,1 dan dan maksimum sebanyak 72 peubah penjelas pada peluang mutasi 0,4. Berdasarkan Gambar 4, rentang panjang gelombang terpilih berada pada 240 nm-295 nm. Panjang gelombang terpilih yang sering banyak muncul berada pada 264 nm dan 272 nm.

Tabel 6: Jumlah panjang gelombang terpilih dengan akurasi 100%

Kombinasi Parameter	Jumlah Peubah Terpilih	Rentang Panjang Gelombang (nm)
0,1;0,1;30;10	40	242,5-292
0,1;0,1;30;5	35	240-295,5
0,3;0,1;30;10	56	240,5-284,5
0,3;0,1;30;5	44	240-287,5
0,5;0,1;30;10	45	240-287,5
0,5;0,1;30;5	38	240-287



Gambar 4: Dotplot panjang gelombang terpilih pada akurasi 100%.

4. Simpulan dan Saran

Berdasarkan hasil penelitian dapat disimpulkan bahwa metode support vector machine (SVM) dengan identifikasi panjang gelombang menggunakan algoritma genetika sangat efektif dalam mengklasifikasikan enam jenis tanaman rimpang jahe emprit, jahe merah, lempuyang emprit, lempuyang gajah, lengkuas merah dan lengkuas dengan memberikan akurasi sebesar 100% pada peluang mutasi 0,1 dan ukuran populasi 30. Parameter ukuran populasi dan parameter peluang mutasi

berpengaruh signifikan terhadap nilai akurasi pada taraf 5%. Selain itu, metode algoritma genetika efektif dalam mereduksi peubah penjelas pada data yang relatif besar.

Daftar Pustaka

- Chen, Q., Zhao, J., Fang, C. and Wang, D. (2007). Feasibility study on identification of green, black and oolong teas using near-infrared reflectance spectroscopy based on support vector machine (svm), *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 66(3): 568-574.
- Chudiwal, A. K., Jain, D. P. and Somani, R. S. (2010). Alpinia Galangal Willd An Overview of Phyto-pharmacological Propertis, *Indian Journal of Natural Products and Resources* 1(2): 143-149.
- Hardianti, H. (2019). Perbandingan Analisis Diskriminan Kanonik dan Diskriminan Kuadrat Terkecil Parsial untuk Klasifikasi Enam Spesies Famili Zingiberaceae [Tesis]. Bogor (ID): Institut Pertanian Bogor
- Khemphila, Boonjing, A. and Veera. (2011). Heart Disease Classification Using Neural Network and Feature Selection, *IEEE Computer Society*.
- Leardi, R. (2001). Genetic Algorithms in Chemometrics and Chemistry: A Review, *Journal of Chemometrics*.
- Sartono, B. (2010). Pengenalan Algoritma Genetik Untuk Pemilihan Peubah Penjelas Dalam Model Regresi Menggunakan SAS/IML. *Forum Statistika dan Komputasi* 15(2): 10-15.
- Steinberg, D., Colla, P. (2009). CART: Tree-Structured Nonparametric Data Analysis. San Diego: Salford System.