

PEMANFAATAN *HIERARCHICAL CLUSTERING* UNTUK PENGELOMPOKKAN DAUN BERDASARKAN FITUR *MOMENT* *INVARIANT*

Febri Liantoni¹, Laili Cahyani²

Institut Teknologi Adhi Tama Surabaya, Indonesia

Universitas Trunojoyo Madura, Indonesia

Email: febri.liantoni@gmail.com, laili.cahyani12@gmail.com

Abstrak

Ilmu mengenai tanaman telah mengalami kemajuan yang pesat. Salah satunya cabang ilmu mengenai morfologi tanaman. Ilmu morfologi ini mempelajari susunan tubuh tanaman khususnya mengenai bentuk tepi daun. Pada penelitian ini akan dilakukan pengelompokkan daun berdasarkan bentuk tepi daun. Metode yang digunakan untuk melakukan pengelompokkan adalah metode *Centroid Linkage Clustering* yang merupakan bagian dari algoritma *Hierarchical Clustering*. Metode ini dikenal lebih memiliki beban komputasi yang relatif lebih ringan karena hanya cukup menghitung titik tengah antar *cluster*. Berdasarkan hasil uji coba yang dilakukan, penggunaan metode *Centroid Linkage Clustering* didapatkan nilai akurasi *clustering* sebesar 87%, sedangkan dengan menggunakan metode k-means didapatkan nilai akurasi *clustering* sebesar 81%. Hal ini menunjukkan bahwa kinerja metode *Centroid Linkage Clustering* lebih baik dibandingkan metode k-means.

Kata Kunci: Morfologi, *Centroid Linkage Clustering*, *Hierarchical Clustering*, *Cluster*, *K-mean*.

Abstract

Science of the plant has progressed rapidly. One of these branches of knowledge about morphology. Science is examining the morphology of plants, especially on the edge of the leaf shape. This research will be conducted grouping leaf by edge shape. The method used to carry out the method of grouping is Centroid Linkage Clustering algorithms that part of the Hierarchical Clustering. This method has the computational load is known to be relatively lighter because it just simply calculate the midpoint between clusters. Based on the results of tests conducted, Centroid Linkage Clustering method haved accuracy values obtained by 87%, while using the k-means clustering method values obtained accuracy of 81%. Performance of the Centroid Linkage Clustering method better than k-means method.

Keywords : *Morphology, Centroid Linkage Clustering, Hierarchical Clustering, Cluster, K-mean.*

PENDAHULUAN

Ilmu Botani atau tentang tumbuhan-tanaman telah mengalami kemajuan yang sangat pesat, bidang-bidang pengetahuan yang sebelumnya hanya merupakan cabang-cabang ilmu tanaman saja, sekarang ini telah menjadi bidang ilmu yang berdiri sendiri-sendiri.

Dari berbagai bidang ilmu tanaman yang sekarang telah berdiri sendiri salah satunya adalah Morfologi Tanaman. Morfologi Tanaman yang mempelajari bentuk dan susunan tubuh tanaman pun sudah demikian pesat perkembangannya hingga dipisahkan menjadi morfologi luar atau morfologi saja dan morfologi dalam atau anatomi tanaman. Salah satunya diambil pembahasan mengenai morfologi tanaman berdasarkan bentuk tepi daun.

Tanaman merupakan bagian paling penting dalam kehidupan di muka bumi. Tanaman berguna sebagai penyedia oksigen untuk bernafas, sebagai bahan makanan, bahan bakar, obat-obatan, kosmetik dan lebih banyak lagi. Proses pengelompokkan tanaman dapat dilakukan dengan cara mengidentifikasi citra bentuk daun dari tanaman itu sendiri. Cara pengambilan gambar daun dari tanaman tersebut, maka dapat dilakukan langkah-langkah pengenalan pola daun dengan cara mengenali karakteristik struktural daun seperti bentuk dan tekstur daun tersebut[1][2][3][4].

Pada proses pengelompokkan tanaman dapat dilakukan dengan cara mengidentifikasi gambar bentuk daun dari tanaman. Dengan cara tersebut maka dapat dilakukan langkah-langkah pengenalan pola daun dengan mengenali karakteristik struktural daun seperti bentuk dan tekstur sebuah daun. Metode untuk melakukan pemrosesan terhadap citra masukan dengan pemanfaatan teknik pengolahan citra digital

dilakukan untuk menganalisa karakteristik struktural daun.

Perkembangan teknologi untuk teknik pengolahan citra juga berkembang pesat. Berbagai teknik dikembangkan untuk mempermudah pekerjaan manusia, baik sebagai pengolah citra, analisis citra maupun pengguna citra untuk berbagai tujuan dan keperluan. Seringkali citra yang digunakan tidak dalam kondisi yang ideal untuk dikaji dikarenakan banyaknya gangguan, dapat berupa bayangan, foto atau gambar kabur, kurang jelasnya kenampakan obyek sehingga dapat menimbulkan masalah dan mempengaruhi hasil interpolasi serta akan mempengaruhi analisa dan perencanaan yang akan dilakukan, maka diperlukan berbagai teknik pengolahan citra untuk memperoleh citra yang ideal.

Teknik pengolahan citra digital ini dilakukan pada tahapan praproses citra sampai didapatkan bentuk tepian dan ciri struktural dari masing-masing daun. Metode yang digunakan pada ekstraksi fitur ini adalah pengenalan fitur morfologi digital [3]. Proses ekstraksi fitur yang digunakan menggunakan metode *Moment Invariant*. Setelah itu dilakukan ekstraksi fitur dari citra daun tersebut sehingga didapatkan informasi struktural daun yang kemudian digunakan sebagai data pengelompokkan.

Beragamnya jenis tanaman dengan berbagai bentuk tepi daun, maka akan cukup sulit untuk melakukan pengelompokkan tanaman berdasarkan morfologi daun. Akan lebih mudah jika terdapat sistem otomatis yang dapat mengelompokkan tanaman sesuai dengan morfologinya. Beberapa contoh penelitian tentang pengelompokkan bentuk tanaman, umumnya menggunakan metode jaringan saraf tiruan. [1][2][3][4]. Metode jaringan saraf tiruan banyak digunakan karena metode ini dikenal lebih cepat secara substansial. Akan tetapi

penentuan jumlah hidden layer yang digunakan berpengaruh pada hasil, serta dibutuhkan parameter jumlah epoch yang besar sehingga membutuhkan komputasi yang lebih tinggi. Pada tahun 2012, Arunpriya melakukan penelitian pengenalan tanaman menggunakan algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (K-NN).[5]. Hasil dari metode K-NN sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan. Metode SVM ini bersifat linear classifier dan secara teoritik hanya dikembangkan untuk permasalahan dua kelas. Pada penelitian sebelumnya telah dilakukan klasifikasi daun herbal menggunakan metode Naïve Bayes Classifier Dan K Nearest Neighbor [6].

Pada penelitian ini akan melakukan pengelompokan citra daun dengan *Centroid Linkage Clustering* berdasarkan fitur bentuk tepi daun. Sebagai pembanding digunakan metode k-means. Metode *Centroid Linkage Clustering* merupakan bagian dari karakteristik *Hierarchical Clustering* [7]. Metode ini dikenal baik untuk kasus dengan *normal data set distribution*. Metode ini menghubungkan titik tengah diantara kluster lainnya. [8]. Sebelum dilakukan tahapan pengelompokan, maka terlebih dahulu dilakukan tahapan praproses citra dan ekstraksi fitur citra tepi daun agar didapatkan nilai masukan yang tepat untuk tahapan klasifikasi spesies daun berdasarkan citra daun.

PERANCANGAN SISTEM

Morfologi Tanaman

Morfologi tanaman merupakan ilmu yang mempelajari bentuk dan susunan tubuh tanaman, dipisahkan menjadi morfologi luar atau morfologi saja dan morfologi dalam atau anatomi tanaman. Salah satunya diambil morfologi tanaman berdasarkan bentuk tepi

daun.

Secara konsep macam-macam pembagian bentuk daun adalah sebagai berikut ini [9].

- Integer (pinggir rata)
- Divisus atau bertoreh (pinggir tidak rata)
- Tepi daun dengan toreh yang merdeka (tidak mempengaruhi bentuk umum)
- Serratus (pinggir bergerigi atau bergergaji)
- Dentatus (pinggir bergigi)
- Creatus (pinggir beringgit)
- Tepi daun dengan toreh tidak merdeka (mempengaruhi bentuk umum).
- Lobatus (pinggir berlekuk menyirip dan menjari)
- Fissus / vidus (pinggir bercanggap menyirip dan menjari).
- Paritus / diversifolia (pinggir berbagi menyirip dan menjari).

Data yang digunakan berupa gambar daun dengan *background* berwarna putih. Contoh gambar yang digunakan seperti ditunjukkan pada Gambar 1.



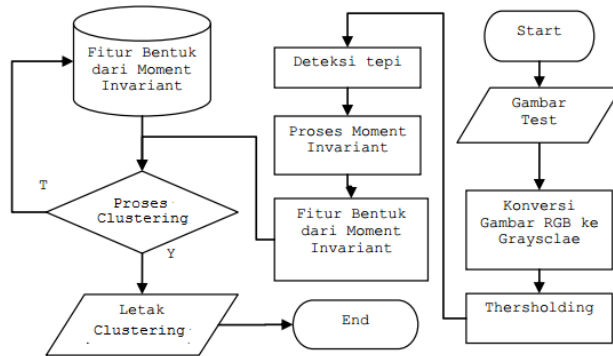
Gambar 1. Contoh data daun

Praproses

Praproses dilakukan dengan tujuan untuk mengolah data masukan sehingga dapat digunakan untuk proses ekstraksi fitur serta

clustering. Terdapat beberapa metode yang digunakan untuk praproses, yaitu perubahan citra pada ruang warna RGB ke ruang warna grayscale, pembuatan citra biner, dan deteksi tepi dengan Laplacian.

Tahapan yang dilakukan pada penelitian ini seperti ditunjukkan pada Gambar 2.



Gambar 2. Tahapan penelitian

Tahap pertama yaitu mengubah citra ke dalam grayscale. Pemrosesan ini dilakukan untuk mengubah domain piksel citra menjadi 8 bit skala abu-abu. Untuk konversi tersebut digunakan seperti ditunjukkan pada Persamaan 1.

$$gray = \frac{red*299+green*587+blue*114}{1000} \quad (1)$$

Kemudian dilakukan pembentukan citra biner dengan menggunakan thresholding. Piksel dengan nilai lebih besar threshold akan dianggap sebagai background sedangkan piksel yang kurang dari threshold akan dianggap sebagai objek daun.

Setelah itu dilakukan deteksi tepi dari citra dengan memeriksa diskontinuitas pada nilai intensitas. Diskontinuitas tersebut dapat dideteksi dengan menggunakan *first-order derivatives* dan *second-order derivatives*. *First-order derivatives* dari fungsi gradient citra 2D

dengan persamaan fungsi $f(x,y)$ ditunjukkan pada Persamaan 2.

$$\nabla f = \begin{bmatrix} Gx \\ Gy \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (2)$$

Second-order derivatives dihitung menggunakan Laplacian dari $f(x,y)$ seperti ditunjukkan Persamaan 3.

$$\nabla^2 f(x, y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \quad (3)$$

Ekstraksi Fitur

Ekstraksi fitur merupakan suatu proses untuk menghasilkan nilai-nilai fitur berupa vektor fitur dari citra biner daun yang telah dilakukan deteksi tepi. Vektor fitur tersebut kemudian digunakan untuk tahap clustering. Fitur yang digunakan yaitu seven moment invariant yang akan menghasilkan tujuh nilai pada vektor fitur.

Moment Invariant

Proses pengenalan sebuah obyek di dalam sebuah citra setelah proses segmentasi, sering terbentur pada permasalahan posisi obyek, rotasi sumbu obyek, dan perubahan skala dari obyek. Posisi obyek yang bergeser atau berputar maupun ukurannya yang lebih kecil atau lebih besar daripada dapat menyebabkan kesalahan dalam pengenalan atau identifikasi obyek.

Pada penggunaan perhitungan nilai 2 (dua) dimensi momen sample gambar $M \times M$ dari fungsi kontinu $f(x,y), (x,y=0, \dots, M-1)$ didapatkan Persamaan 4.

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} (x)^p \cdot (y)^q \cdot f(x, y) \quad (4)$$

Momen dapat menggambarkan suatu obyek dalam hal area, posisi, orientasi dan parameter terdefinisi lainnya. Dengan

mendapatkan sejumlah informasi momen, baik momen tingkat ke nol (m_{00}) dan kesatu (m_{10} dan m_{01}) atau momen sentral, dan momen pada tingkat ≥ 2 atau moment invariant dari sebuah obyek, maka obyek tersebut dapat diidentifikasi sekalipun telah mengalami pergeseran (translasi), perputaran (rotasi) maupun perubahan skala.

Dari moment $f(x,y)$ akan ditranslasikan dengan nilai (a,b) sehingga didapatkan perhitungan baru seperti Persamaan 5.

$$\mu_{pq} = \sum_x \sum_y (x + a)^p \cdot (y + b)^q \cdot f(x, y) \tag{5}$$

Dari central moment utama yaitu m_{pq} atau μ_{pq} dikomputasi melalui proses substitusi terhadap nilai $a = -x$ dan nilai $b = -y$ maka akan didapatkan perhitungan pada Persamaan 6.

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p \cdot (y - \bar{y})^q f(x, y) \tag{6}$$

$$\text{Untuk nilai } \bar{x} = \frac{m_{10}}{m_{00}} \text{ dan } \bar{y} = \frac{m_{01}}{m_{00}}$$

Ketika proses normalisasi maka nilai penskalaan yang digunakan dalam perhitungan pada Persamaan 7.

$$\eta_{pq} = \mu_{pq} / \mu_{pq}^\gamma \tag{7}$$

$$\text{Untuk nilai } \gamma = [(p + q)/2] + 1$$

Kemudian dari proses tersebut maka didapatkan nilai *seven moment invariant* dengan perhitungan pada Persamaan 8.

$$\begin{aligned} Hu_1 &= \eta_{20} + \eta_{02} \\ Hu_2 &= (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \\ Hu_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ Hu_4 &= (\eta_{30} + \eta_{12})^2 + (3\eta_{21} + \eta_{03})^2 \end{aligned}$$

$$Hu_5 = (\eta_{30} + 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3\eta_{21} + \eta_{03}]$$

$$Hu_6 = (\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})^2 - 4\eta_{11}\eta_{30} + \eta_{12}\eta_{21} + \eta_{03}$$

$$Hu_7 = (3\eta_{21} - \eta_{30})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3\eta_{21} + \eta_{03}] + \eta_{12}\eta_{21} + \eta_{03} \tag{8}$$

Normalisasi Data

Normalisasi data merupakan sebuah metode untuk menyamakan interval dari semua nilai variabel. Normalisasi dilakukan ketika terdapat variabel yang memiliki kisaran nilai sangat besar, sedangkan variabel lainnya memiliki kisaran nilai sangat kecil. Nilai yang telah dinormalisasi akan memiliki interval antara 0 dan 1.

Normalisasi dilakukan dengan menggunakan Persamaan 9 [2].

$$norm = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{9}$$

Dimana X adalah nilai variabel data, X_{min} adalah nilai terkecil pada variabel tersebut, dan X_{max} adalah nilai terbesar pada variabel tersebut.

Pengelompokan Data (Clustering)

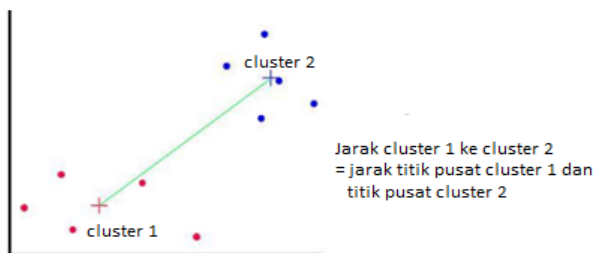
Clustering adalah suatu metode untuk mengelompokkan data yang memiliki kesamaan ke dalam suatu *cluster* tertentu. Sedangkan data yang berada pada *cluster* yang berbeda akan memiliki sedikit kesamaan. Salah satu algoritma clustering adalah *Hierarchical Clustering*. Algoritma ini dimulai dengan menganggap semua data adalah sebuah *cluster*. Pada iterasi berikutnya, *cluster* akan berkurang secara iteratif dan *cluster* yang memiliki jarak terdekat akan

digabungkan, hingga akhirnya iterasi berhenti ketika jumlah *cluster* sesuai dengan jumlah yang diinginkan. Salah satu variasi dari *Hierarchical Clustering* adalah *Centroid Linkage Clustering*.

Selain menggunakan metode *Centroid Linkage Clustering* penelitian ini juga melakukan pengujian dengan menggunakan metode *k-means*. Proses ini dilakukan sebagai langkah pembanding terhadap nilai hasil dari metode *clustering* yang digunakan.

Centroid Linkage Clustering

Centroid Linkage Clustering adalah proses *clustering* yang didasarkan pada jarak antar centroid-nya. Metode ini baik untuk memperkecil *variance within cluster* karena melibatkan centroid pada saat penggabungan antar cluster. Ilustrasi proses metode *Centroid Linkage Clustering* seperti ditunjukkan pada Gambar 3.



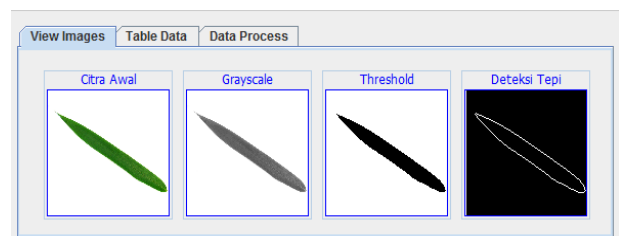
Gambar 3. Ilustrasi metode *centroid linkage clustering*

HASIL DAN PEMBAHASAN

Sistem dibuat menggunakan bahasa pemrograman Java dengan aplikasi *Integrated Development Environment (IDE) Netbeans*. Data citra daun yang digunakan diperoleh dari dataset *Flavia* dan hasil foto. Dari dataset tersebut, digunakan 10 jenis daun dengan jumlah masing-masing jenis adalah 10 buah. Sehingga total data yang digunakan berjumlah 100 data citra daun. Pengujian dilakukan dengan menghitung jumlah data yang telah berada pada cluster yang sesuai.

Hasil Praproses

Dari hasil uji coba sistem yang dilakukan, sistem telah berhasil melakukan praproses yakni sistem telah mengubah gambar dalam ruang warna RGB ke dalam grayscale, kemudian menghasilkan citra biner dengan menggunakan metode *thresholding*, serta menghasilkan deteksi tepi menggunakan metode Laplacian. Contoh hasil seperti ditunjukkan pada Gambar 4.



Gambar 4. Hasil praproses

Hasil yang diperoleh dari praproses seperti yang ditunjukkan pada Gambar 4 meliputi grayscale, threshold, dan deteksi tepi. Setelah didapatkan deteksi tepi dari citra daun kemudian dilakukan ekstraksi fitur untuk proses selanjutnya.

Hasil Ekstraksi Fitur

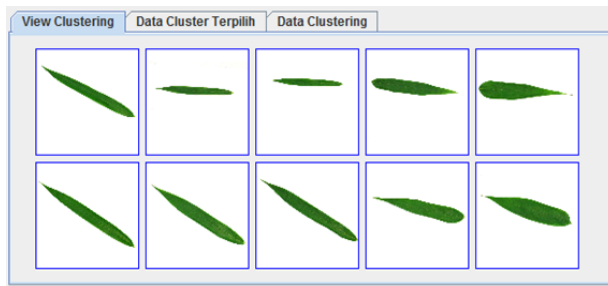
Tahap ekstraksi fitur telah berhasil menghasilkan vektor fitur yang berisi tujuh nilai dari *seven moment invariant*. Data hasil proses ekstraksi kemudian dilakukan proses normalisasi pada interval antara 0 dan 1. Nilai pada vektor fitur hasil normalisasi yang ditampilkan oleh sistem seperti ditunjukkan pada Gambar 5.

1	2	3	4	5	6	7	NamaFile
0.8756097	0.8129986	0.774646	0.7722838	0.59758	0.6968731	0.4025437	D:\Modul ...
0.9026957	0.8588586	0.8265435	0.8240294	0.6802701	0.7641147	0.3198732	D:\Modul ...
0.8769526	0.8179285	0.7802192	0.7778764	0.6062448	0.7040373	0.3938757	D:\Modul ...
0.8799414	0.8229103	0.7858397	0.7834728	0.6149952	0.7112447	0.3851272	D:\Modul ...
0.8578619	0.7840722	0.7421898	0.7399648	0.5486357	0.6558066	0.4514733	D:\Modul ...
0.8710146	0.8080763	0.7691282	0.7668267	0.5891545	0.6898705	0.4109636	D:\Modul ...
0.890441	0.8432025	0.8088153	0.8064023	0.6514804	0.7409776	0.3486471	D:\Modul ...

Gambar 5. Hasil data normalisasi fitur *moment invariant*

Hasil Clustering

Setelah dipilih salah satu metode *clustering*, yang akan digunakan, maka sistem akan menampilkan gambar dari beberapa citra yang berada pada cluster yang sama. Contoh hasil uji coba menggunakan metode Centroid Linkage Clustering seperti ditunjukkan pada Gambar 6.



Gambar 6. Hasil uji coba metode centroid linkage clustering

Hasil Perhitungan Pengujian

Hasil pengujian untuk mengukur kualitas *clustering* akan ditampilkan dengan menunjukkan jumlah anggota yang terdeteksi dengan benar yang terdapat pada sebuah *cluster*. Berikut ini merupakan jumlah anggota benar dari sepuluh *cluster* yang ada untuk algoritma *Centroid Linkage Clustering*. Hasil ditunjukkan pada Tabel 1.

Tabel 1 Data hasil *centroid linkage clustering*

Centroid Linkage Clustering	
Nomor Cluster	Jumlah data benar
1	10
2	10

3	5
4	9
5	9
6	7
7	10
8	8
9	10
10	9
Total	87

Hasil pengujian dari dataset sebanyak 100 data dengan menggunakan metode *Centroid Linkage Clustering* dihasilkan 87 data yang dapat terkelompok secara sempurna. Sehingga dapat disimpulkan bahwa akurasi dari proses clusteringnya adalah $87/100 = 87\%$.

Sebagai pembandingan dilakukan uji coba dengan menggunakan metode k-means. Hasil uji coba dengan menggunakan metode k-means ditunjukkan pada Tabel 2.

Tabel 2 Data hasil k-means

K-means	
Nomor Cluster	Jumlah data benar
1	7
2	10
3	10
4	10
5	10
6	5
7	5
8	6
9	10
10	8
Total	81

Hasil pengujian dari dataset sebanyak 100 data dengan menggunakan metode k-means dihasilkan 81 data yang dapat terkelompok secara sempurna. Sehingga dapat disimpulkan bahwa akurasi dari proses clusteringnya adalah $81/100 = 81\%$.

KESIMPULAN

Dari hasil uji coba yang telah dilakukan meliputi proses praproses, ekstraksi fitur dan clustering dapat diambil kesimpulan antara lain.

- 1) Pada penggunaan metode *seven moment invariant* dapat dibuktikan bahwa perubahan posisi, rotasi dan skala tidak begitu berpengaruh pada nilai fitur invarian yang dihasilkannya.
- 2) Pada uji coba dengan menggunakan metode *centroid linkage clustering* didapatkan nilai akurasi *clustering* sebesar 87%, sedangkan dengan menggunakan metode k-means didapatkan nilai akurasi *clustering* sebesar 81%. Hal ini menunjukkan bahwa kinerja metode *centroid linkage clustering* lebih baik dibandingkan metode k-means.

DAFTAR PUSTAKA

- [1] Stephen G. W., Forrest S. B., Eric Y. Xu, Yu-Xuan W., Yi-F. C. and Qiao-Liang X., 2007, *A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network*, IEEE International Symposium, pp 11-16, July.
- [2] A. Kadir, Lukito E. N, Adhi N, 2011, *Leaf Classification Using Shape, Color, and Texture Features*, International Journal of Computer Trends and Technology, July to Aug.
- [3] Z. Husin, A. Y. M. Shakaff, A. H. A. Aziz, R. S. M. Farook, M. N. Jaafar, U. Hashim, A. Harun, 2012, *Embedded Portable Device For Herb Leaves Recognition Using Image Processing Techniques And Neural Network Algorithm*, Science Direct on Computers and Electronics in Agriculture, pp 18–29.
- [4] Chaki J, Parekh R, 2011, *Plant Leaf Recognition using Shape based Features and Neural Network Classifiers*, International Journal of Advanced Computer Science and Applications, vol 2, no 10.
- [5] ArunPriya C, Balasaravanan T, 2012, *An Efficient Leaf Recognition Algorithm for Plant Classification Using Support Vector Machine*, Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, March 21-23.
- [6] Liantoni, F., Nugroho, H., 2015, *Klasifikasi Daun Herbal Menggunakan Metode Naïve Bayes Classifier Dan Knearest Neighbor*, Jurnal SimanteC, vol 5, no 1, hal 9 – 16.
- [7] Székely, G. J. and Rizzo, M. L., 2005, *Hierarchical clustering via Joint Between-Within Distances: Extending Ward's Minimum Variance Method*, Journal of Classification, pp 151-183.
- [8] Bien J, Tibshirani R, 2011, *Hierarchical Clustering With Prototypes via Minimax Linkage*, Journal of the American Statistical Association
- [9] Gembong, T, 2005, *Morfologi Tanaman*, Gadjah Mada University, Yogyakarta.