

Pengelompokan Kayu Kelapa Menggunakan Algoritma *K-Means* Berdasarkan Tekstur Citra Kayu Kelapa Dua Dimensi (2D)

Setiawan Qodri Nugroho¹, Ricardus Anggi Pramunendar²
 Fakultas Ilmu Komputer, Teknik Informatika, Universitas Dian Nuswantoro^{1,2}
 Email : setiawanqodrinugroho@gmail.com¹, ricardus.anggi@dsn.dinus.ac.id²

Abstract: Generally, the process of selecting coconut wood as the raw material of a product carried out by a trained person (grader) by observing the coconut wood directly without using any tools (manual). This causes the dependence of expertise and experience of a grader in the selection of raw materials coconut wood. The dependency causes problems increasingly complex when data coconut wood to be tested manually increase more and more. This research aims to create and build clustering of coconut wood automatically by implementing the texture feature extraction method based on histogram as a determinant of coconut wood 2D image characteristics and K-Means clustering method to clustering the extracted texture features of coconut wood 2D image. The results showed from 105 data test of coconut wood 2D image labeled A, B, and C successfully grouped into 3 groups with f-measure 93%.

Keywords: *Computer Vision, Coconut Wood Clustering, K-Means Clustering, Statistical Histogram.*

I. PENDAHULUAN

Pohon kelapa tumbuh menyebar dari pantai hingga pegunungan terutama di daerah tropis dan pohon ini dapat ditemukan tumbuh secara liar atau ditanam sebagai usaha masyarakat seperti di Indonesia [1]. Indonesia merupakan negara kedua terbesar yang memiliki pertanaman kelapa setelah negara Filipina. Tanaman ini tumbuh menyebar dan dapat dijumpai hampir diseluruh kepulauan Indonesia [2].

Penentuan penilaian kelas secara visual pada kayu kelapa ditentukan oleh kepadatan bundel (pola) pada irisan kayu kelapa per satuan luas. Semakin padat maka kualitas kayu semakin baik. Hal lain yang membuat faktor kualitas kayu kelapa adalah warna dan ukuran pola bundel, ukuran yang lebih besar dan warna lebih gelap akan menghasilkan kualitas yang lebih baik [4].

Kebutuhan akan kayu kelapa yang berkualitas untuk produk yang bermutu, pada industri mebel diperlukan kontrol yang baik pada semua proses, mulai dari pemilihan bahan baku, pengolahan sampai dengan produk tersebut jadi. Pada umumnya proses pemilihan bahan baku kayu kelapa dilakukan oleh seorang pakar (*grader*) secara konvensional, hal tersebut menyebabkan ketergantungan akan keahlian dan pengalaman seorang *grader* dalam melakukan pemilihan bahan baku kayu kelapa [5]. Ketergantungan tersebut menyebabkan masalah semakin kompleks ketika jumlah kayu kelapa yang akan diuji secara visual semakin banyak.

Pada penelitian ini digunakan ekstraksi fitur tekstur berbasis *histogram* untuk mendapatkan 6 karakteristi ciri tekstur citra kayu kelapa dan metode *clustering* dengan algoritma *K-Means* yang digunakan untuk mengelompokkan data data hasil ekstraksi fitur tekstur berbasis histogram yang banyak jumlahnya ke dalam 3

golongan kelompok. Dari tiga kelompok kayu kelapa yang telah berhasil dikelompokkan selanjutnya akan dibandingkan dengan data uji kualitas kayu kelapa dari Pendidikan Industri Kayu (PIKA) yang didapatkan dari peneliti sebelumnya dengan penelitian [4] [5]. Hal ini dilakukan untuk menganalisa kinerja serta menguji presisi, *recall* dan *f-measure* dari algoritma *K-Means clustering* dalam mengelompokkan kayu kelapa dengan metode ekstraksi fitur tekstur berbasis *histogram*.

II. METODE YANG DIUSULKAN

Peneliti menerapkan beberapa metode yang bertujuan untuk menyelesaikan masalah yang ada dalam penelitian ini. Metode yang diusulkan dalam penelitian ini adalah sebagai berikut :

A. Ekstraksi Fitur Tekstur Berbasis Histogram

Merupakan suatu metode untuk mendapatkan tekstur dengan mendasarkan pada *histogram*. *Histogram* citra adalah grafik yang menggambarkan penyebaran nilai-nilai intensitas piksel dari suatu citra atau bagian tertentu di dalam citra. Dari sebuah *histogram* dapat diketahui frekuensi kemunculan nisbi (*relative*) dari intensitas pada citra tersebut [13]. Metode *histogram* merupakan metode statis orde satu untuk memperoleh fitur tekstur. Fitur-fitur yang dapat dikenali melalui metode *histogram* diantaranya: rerata intensitas, deviasi standar, *skewness*, energi, entropi dan kehalusan [10]. Persamaan penghitungan ke-6 ciri fitur tekstur berbasis histogram dapat dilihat dibawah ini [10] [12] [21]:

Fitur pertama yang dihitung secara statistis adalah rerata intensitas. Komponen fitur ini dihitung dengan menggunakan persamaan sebagai berikut :

$$m = \sum_{i=0}^{L-1} i \cdot p(i) \quad (1)$$

Dalam hal ini, i adalah aras keabuan pada citra f dan $p(i)$ menyatakan probabilitas kemunculan i dan L menyatakan nilai aras keabuan tertinggi. Rumus pada persamaan (1) akan menghasilkan rerata intensitas objek [10].

Fitur kedua berupa deviasi standar. Perhitungannya sebagai berikut :

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (i - m)^2 p(i)} \quad (2)$$

Dalam hal ini, σ^2 dinamakan varians atau momen orde dua ternormalisasi karena $p(i)$ merupakan fungsi peluang. Fitur ini memberikan ukuran kekонтarasan [10].

Fitur *skewness* merupakan ukuran ketidaksimetrisan terhadap rerata intensitas. Berikut persamaannya :

$$skewness = \sum_{i=0}^{L-1} (i - m)^3 p(i) \quad (3)$$

Skewness sering disebut sebagai momen orde tiga ternormalisasi. Nilai negatif menyatakan bahwa distribusi kecerahan condong ke kiri terhadap rerata dan nilai positif menyatakan bahwa distribusi kecerahan condong ke kanan terhadap rerata. Dalam praktik, nilai *skewness* dibagi dengan $(L-1)^2$ supaya ternormalisasi [10].

Deskriptor energi adalah ukuran yang menyatakan distribusi intensitas piksel terhadap jangkauan aras keabuan. Definisinya sebagai berikut :

$$energi = \sum_{i=0}^{L-1} [p(i)]^2 \quad (4)$$

Citra yang seragam dengan satu nilai aras keabuan akan memiliki nilai energi yang maksimum, yaitu sebesar 1. Secara umum, citra dengan sedikit aras keabuan akan memiliki energi yang lebih tinggi daripada yang memiliki banyak nilai aras keabuan. Energi sering disebut sebagai keseragaman [10].

Entropi mengindikasikan kompleksitas citra. Perhitungannya sebagai berikut :

$$entropi = - \sum_{i=0}^{L-1} p(i) \log_2(p(i)) \quad (5)$$

Semakin tinggi nilai entropi, semakin kompleks citra tersebut. Perlu diketahui, entropi dan energi berkecenderungan berkebalikan. Entropi juga merepresentasikan jumlah informasi yang terkandung di dalam sebaran data [10].

Properti kehalusan biasa disertakan untuk mengukur tingkat kehalusan/kekasaran intensitas pada citra. Persamaannya sebagai berikut :

$$R = 1 - \frac{1}{1 + \sigma^2} \quad (6)$$

Pada rumus persamaan diatas, σ adalah deviasi standar. Nilai R yang rendah menunjukkan bahwa citra memiliki intensitas yang kasar. Perlu diketahui, di dalam menghitung kehalusan, varians perlu dinormalisasi sehingga nilainya berada dalam jangkauan [0 1] dengan cara membaginya dengan $(L-1)^2$ [10].

B. K-Means Clustering

K-Means merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data ke dalam *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari *clustering* ini adalah untuk meminimalisasi *objective function* yang diset dalam proses *clustering*, yang ada umumnya berusaha meminimalisasi variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* [7].

Metode *K-Means Cluster Analysis* menggunakan algoritma sebagai berikut [7]:

1. Menentukan k sebagai jumlah *cluster* atau jumlah kelompok yang ingin di bentuk.
2. Membangkitkan k *centroid* (titik pusat *cluster*) awal secara *random*. Penentuan *centroid* awal dilakukan secara *random* atau acak dari objek-objek yang tersedia sebanyak k *cluster*.
3. Menghitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *cluster*. Untuk menghitung jarak antara objek dengan *centroid*, menggunakan rumus *Euclidian Distance* :

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (7)$$

Dimana x_i : objek x ke- i

y_i : daya y ke- i

n : banyaknya objek

4. Mengalokasikan masing-masing objek ke dalam *centroid* yang paling terdekat. Untuk melakukan pengalokasian objek ke dalam masing-masing *cluster* pada saat iterasi secara umum dapat dilakukan dengan menyatakan setiap objek sebagai anggota *cluster* dengan mengukur jarak kedekatan sifatnya terhadap titik pusat *cluster* tersebut.
5. Melakukan iterasi kemudian menentukan posisi *centroid* baru dengan menggunakan persamaan :

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad (8)$$

Dimana v : *centroid* pada *cluster*

x_i : objek ke- i

n : banyaknya objek / jumlah objek yang menjadi anggota *cluster*

6. Mengulangi ke langkah 3 jika posisi *centroid* baru tidak konvergen. Pengecekan *konvergensi* dilakukan dengan membandingkan matriks *group assignment* pada iterasi sebelumnya dengan matriks *group assignment* pada iterasi yang sedang berjalan. Jika hasilnya sama maka algoritma *K-Means cluster* dinyatakan sudah konvergen dan proses berhenti. Tetapi jika belum konvergen maka perlu mengulangi iterasi ke langkah 3.

C. Evaluation Measure

Penelitian ini menggunakan teknik *F-Measure* untuk mengukur kinerja yang diusulkan. Pengukuran *F-Measure* berdasarkan pada nilai pengukuran presisi dan *recall* antara *class* hasil *cluster* dengan *class* sebenarnya yang terdapat pada data masukan [22]. Semakin tinggi nilai *F-Measure* maka menunjukkan tingkat akurasi yang tinggi dari hasil *clustering* [23]. *Recall* dihitung untuk mengevaluasi seberapa besar cakupan suatu model *clustering* dalam memprediksi suatu kelas tertentu sedangkan presisi dihitung untuk mengevaluasi seberapa baik ketepatan model dapat memprediksi suatu kelas [24]. Nilai presisi dan *recall* kategori i dalam *cluster* j dapat diperoleh dari persamaan berikut [23] :

$$\text{Presisi}(i, j) = \frac{n_{ij}}{n_j} \quad (9)$$

$$\text{Recall}(i, j) = \frac{n_{ij}}{n_i} \quad (10)$$

Dimana :

n_{ij} : Jumlah data kategori i dalam *cluster* j

n_i : Jumlah data dalam kategori i

n_j : Jumlah data dalam *cluster* j

Penghitungan *F-Measure* menggunakan persamaan berikut [23] :

$$F(i, j) = \frac{2 * (\text{Presisi} * \text{Recall})}{(\text{Presisi} + \text{Recall})} \quad (11)$$

Rata – rata perhitungan *F-Measure* menggunakan persamaan berikut [23] :

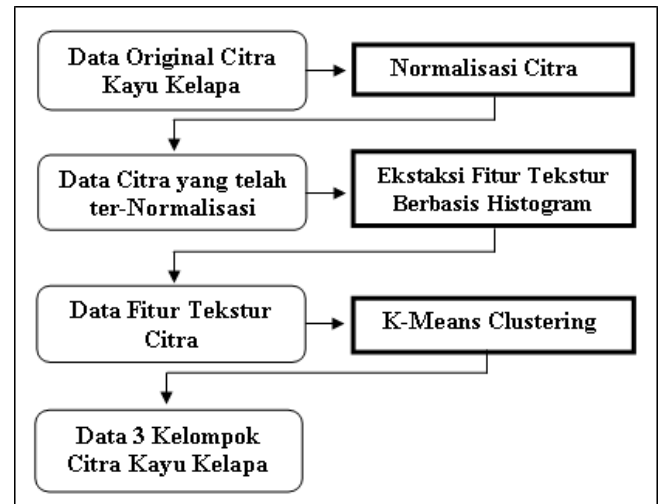
$$F = \sum_i \frac{n_i}{n} \text{Max}_{j=1, \dots, k} F(i, j) \quad (12)$$

Dimana :

$\text{Max}\{F(i, j)\}$ merupakan nilai *max F-Measure* dari kategori i dalam *cluster* j

III. IMPLEMENTASI

Peneliti menerapkan beberapa prosedur penyelesaian yang bertujuan untuk menyelesaikan masalah yang ada dalam penelitian ini. Prosedur penyelesaian yang diusulkan dalam penelitian ini adalah sebagai berikut :



Gambar 1: Metode Penyelesaian

Beberapa tahap implementasi yang dilakukan dalam penelitian ini adalah sebagai berikut :

A. Persiapan data citra kayu kelapa 2D

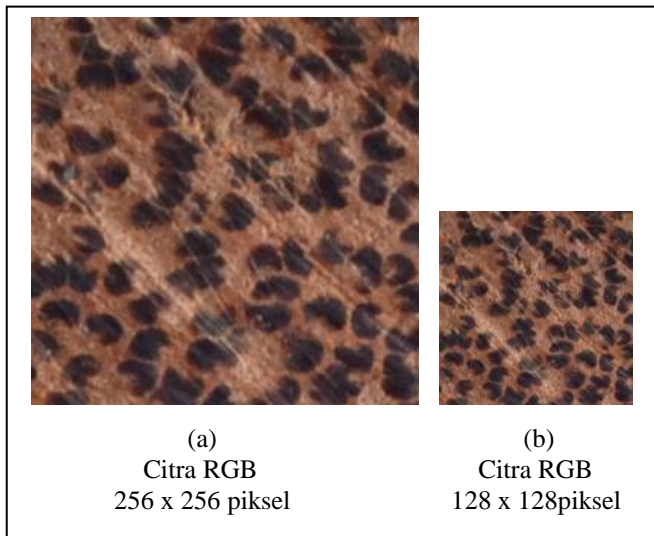
Dalam melakukan eksperimen ini, diperlukan citra digital 2D dari kayu kelapa dengan potongan melintang. Data yang digunakan dalam penelitian ini berjumlah 105 data, dan data penelitian tersebut didapatkan dari penelitian sebelumnya [4] [5]. Pada data tersebut sebelumnya telah dilakukan proses pensortiran oleh *grader* untuk diklasifikasikan kelas kualitasnya.



Gambar 2: Contoh data kayu kelapa yang diuji oleh *grader* dengan kualitas A

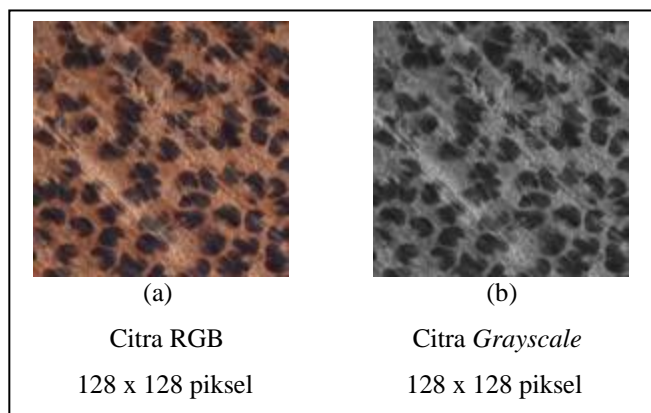
B. Normalisasi Citra Kayu Kelapa

Data citra kayu kelapa merupakan citra RGB yang memiliki dimensi ukuran resolusi yang besar, sehingga perlu dilakukan normalisasi ke dalam ukuran tertentu. Hal ini bertujuan untuk menyamakan semua citra yang ingin diolah dan meminimalkan konsumsi memori. Dari data citra yang sudah ada, hal pertama yang dilakukan adalah merubah ukuran resolusi data menjadi 128 x 128 piksel.



Gambar 3: Contoh normalisasi citra kayu kelapa ke RGB 128 x 128 piksel

Setelah melakukan proses normalisasi ukuran citra kayu kelapa 2D menjadi 128 x 128 piksel, proses selanjutnya adalah dengan melakukan normalisasi kedua untuk citra kayu kelapa, yaitu dari citra kayu kelapa dengan model RGB dikonversi menjadi bentuk citra berskala keabuan atau *grayscale*. Hal ini dikarenakan dalam proses ekstraksi fitur tekstur berbasis *histogram* terdapat nilai intensitas aras keabuan, dimana nilai intensitas tersebut merupakan nilai penentu dari ke-6 ciri ekstraksi fitur tekstur berbasis *histogram*.



Gambar 3: Contoh normalisasi citra kayu kelapa RGB menjadi grayscale

C. Ekstraksi Fitur Tekstur Citra Kayu Kelapa

Ekstraksi fitur tekstur bertujuan untuk mendapatkan ciri dari tekstur citra kayu kelapa 2D yang sebelumnya telah melalui tahap pengolahan awal citra. Ekstraksi fitur tekstur tersebut menggunakan metode ekstraksi fitur tekstur berbasis *histogram*. Fitur yang didapatkan dari metode ekstraksi fitur tekstur berbasis *histogram* diantaranya rerata intensitas, deviasi standar, *skewness*, energi, entropi dan kehalusan. atau *smoothness*.

Citra	Fitur	
	Rerata intensitas	: 110.9113
	Rerata kontras	: 31.729
	<i>Skewness</i>	: -0.5154
	Energi	: 0.0152
	Entropi	: 4.4876
	<i>Smoothness</i>	: 0.0152

Gambar 3: Contoh ekstraksi fitur tekstur citra kayu kelapa

D. Pengelompokan Data Ekstraksi Citra Kayu Kelapa

Setelah menyelesaikan langkah ekstraksi fitur tekstur berbasis *histogram* dan mendapatkan data fitur tekstur dari citra kayu kelapa 2D, proses selanjutnya adalah melakukan pengelompokan dari data tersebut dengan menggunakan algoritma *K-Means Clustering* dan penghitungan *Euclidian Distance*.

Hal pertama yang dilakukan adalah menentukan banyaknya *cluster* atau banyaknya kelompok yang diinginkan. Pada penelitian ini telah ditentukan bahwa terdapat 3 kelompok yang akan dikelompokkan. Dimana masing – masing nilai *centroid* telah ditentukan berdasarkan data citra kayu kelapa yang sebelumnya sudah dilabelkan oleh *grader* berdasarkan kualitas kayu kelapa, yaitu kelas A, B dan C. Data 3 *centroid* yang digunakan pada penelitian ini adalah data dengan kode kayu C-1, B-1, dan A-1. Selanjutnya akan dihitung nilai jarak antara data ciri fitur tekstur kayu kelapa dengan data *centroid* yang telah dipilih menggunakan *Euclidian Distance*. Proses ini akan terus diulang sampai tidak ada objek yang berubah dalam suatu *cluster* atau suatu kelompok citra kayu kelapa.

IV. HASIL & PEMBAHASAN

Dari 105 data citra kayu kelapa yang telah berhasil dikelompokkan oleh sistem, akan diuji dengan menggunakan *Evaluation Measure*, dimana akan dihitung nilai presisi, *recall* dan *f-measure*. Hasil pengujian dan penghitungan menggunakan *evaluation measure* pada citra uji label A, label B, dan label C terhadap kelompok 1, 2, dan 3 didapatkan nilai presisi, *recall* dan *f-measure* yang dapat dilihat pada tabel 1 :

Tabel 1: Tabel hasil penghitungan *f-measure* per kelompok

F	1	2	3
A	1	0	0
B	0	0.90	0.11
C	0	0.05	0.91

Dari hasil penghitungan *f-measure* yang sebelumnya telah dilakukan dan terdapat pada tabel 1 dapat dihitung nilai rerata perhitungan *f-measure* pada semua kelompok menggunakan persamaan (12) sebagai berikut :

$$\begin{aligned}
 F &= \left(\frac{35}{105} * \text{Max}((A, 1), (A, 2), (A, 3)) \right) + \\
 &\left(\frac{35}{105} * \text{Max}((B, 1), (B, 2), (B, 3)) \right) + \\
 &\left(\frac{35}{105} * \text{Max}((C, 1), (C, 2), (C, 3)) \right) \\
 F &= \left(\frac{35}{105} * \text{Max}((1), (0), (0)) \right) + \\
 &\left(\frac{35}{105} * \text{Max}((0), (0.90), (0.11)) \right) + \\
 &\left(\frac{35}{105} * \text{Max}((0), (0.05), (0.91)) \right) \\
 F &= \left(\frac{35}{105} * 1 \right) + \left(\frac{35}{105} * 0.90 \right) + \left(\frac{35}{105} * 0.91 \right) \\
 &= 0.33 + 0.3 + 0.30 = 0.93
 \end{aligned}$$

Dari hasil penghitungan rerata *f-measure* diatas jika dibuat persentase maka hasilnya adalah sebagai berikut :

$$F = 0.93 \times 100\% = 93\%$$

Dengan hasil penghitungan persentase diatas dapat diketahui bahwa nilai *f-measure* dari pengelompokan yang dilakukan oleh sistem terhadap 105 data uji citra kayu kelapa adalah 93%.

V. PENUTUP

A. Kesimpulan

Kesimpulan yang diperoleh setelah melakukan penelitian adalah sebagai berikut:

1. Dari hasil pengujian terhadap data pengelompokan yang dilakukan sistem terhadap 105 data citra uji kayu kelapa dengan label A, B, dan C didapatkan nilai *f-measure* atau akurasi sebesar 0.93, atau 93%.
2. Dari hasil penelitian dan pengujian yang telah dilakukan, maka metode yang diusulkan pada penelitian ini dapat digunakan untuk mengelompokkan citra kayu kelapa 2D ke dalam 3 kelompok secara otomatis. Sehingga dapat memudahkan seorang *grader* dalam mengelompokkan citra kayu kelapa ke dalam 3 kelompok dimana masing – masing kelompok mewakili kelas kualitas dari kayu kelapa.
3. Warna gelap dan kepadatan pola bundel dari citra kayu 2D sangat mempengaruhi hasil pengelompokkan kayu kelapa. Hal inilah yang menyebabkan perbedaan kelompok pada citra uji kayu kelapa dengan label B dan C.

B. Saran

Berikut merupakan beberapa hal yang perlu diperhatikan untuk melakukan penelitian lebih lanjut :

1. Penelitian ini dapat dilanjutkan dengan beberapa teknik peningkatan citra seperti dengan adanya suatu metode untuk menghilangkan *noise* pada citra. Hal ini dimaksudkan agar pada penelitian selanjutnya nilai akurasi pengujian dapat lebih baik.
2. Penelitian ini dapat dilanjutkan dengan menggunakan metode *clustering* lainnya seperti *K-Medoid*, *Fuzzy C-Means* dan metode *clustering* lainnya, agar dapat diketahui perbedaan hasil penelitian dari metode *clustering* yang telah diimplementasikan pada objek citra kayu kelapa.
3. Pada penelitian selanjutnya diharapkan terdapat *dataset* yang lebih banyak.

DAFTAR PUSTAKA

- [1] I. Y. Wardhani, S. Surjokusumo, Y. S. Hadi, and N. Nugroho, "Sifat Mekanis Kayu Kelapa Bagian dalam Terpadatkan," *RIMBA Kalimantan Falkutas Kehutanan Unmul*, vol. 11, pp. 62-68, Dec. 2003.
- [2] Barly, "Coconut Stem As Conventional Timber Alternative," *Duta Rimba*, vol. 174, pp. 43-53, Dec. 1994.
- [3] I. N. Yuwono, R. A. Premunendar, P. N. Andono, and R. A. Subandi, "The Quality Determination Of Coconut Wood Density Using Learning Vector Quantization," *Journal Of Theoretical and Applied Information Technology*, vol. 57, pp. 82-87, Nov. 2013.
- [4] R. A. Premunendar and C. Supriyanto, "Klasifikasi Kualitas Kayu Kelapa Menggunakan Gray-Level Co-Occurrence Matrix Berbasis Backpropagation dan Algoritma Genetika," *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2014*, pp. 250-253, Nov. 2014.
- [5] A. Kadir and A. Susanto, *Pengolahan Citra Teori dan Aplikasi*. Yogyakarta, Indonesia: Andi, 2013.
- [6] S. Sergy'an, "Color Histogram Features Based Image Classification in Content-Based Image Retrieval System," *International Symposium on Applied Machine Intelligence and Informatics*, pp. 221-224, 2008.
- [7] R. Munir, *Pengolahan Citra Digital*. Bandung, Indonesia: Informatika, 2004.
- [8] H. S. Sheshadri and A. Kandaswamy, "Breast Tissue Classification Using Statistical Feature Extraction Of Mammograms," *Departement of Electronics & Communication Enggining*, vol. 23, pp. 105-107, 2006.
- [9] I. M. K. Wicaksana and I. M. Widiartha, "Penerapan Metode Ant Colony Optimization Pada Metode K-Harmonic Means Untuk Klasterisasi Data," *Jurnal Ilmu Komputer*, vol. 5, pp. 55-61, Apr. 2012.
- [10] A. Lutfiarta, J. Zeniarja, and A. Salam, "Algoritma Latent Symantic Analysis (LSA) Pada Peringkat Dokumen Otomatis Untuk Proses Clustering Dokumen," *Seminar Teknologi Informasi & Komunikasi Terapan*, pp. 13-18, Nov. 2013.
- [11] A. S. Sastrawan, Z. A. Baizal, and M. A. Bijaksana, "Analisis Pengaruh Metode Combine Sampling Dalam Churn Prediction Untuk Perusahaan Telekomunikasi," *Seminar Nasional Informatika 2010*, vol. A, pp. 14-22, 2010.