ARTIKEL

SISTEM PENGENALAN JENIS MANGGA BERDASARKAN DAUN DENGAN METODE K-NEAREST NAIGHBOR

Oleh:

ARIF AGUS TRIONO
13.1.03.02.0224

Dibimbing oleh :

1. **Ardi Sajaya, M.Kom**
2. **Julian Sahertian, S.Pd., M.T**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

2017

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2017

Yang bertanda tangan di bawah ini:

Nama Lengkap	: Arif Agus Triono
NPM	: 13.1.03.02.0224
Telepun/HP	: 082229444077
Alamat Surel (Email)	: arifagustriono93@gmail.com
Judul Artikel	: Sistem Pengenalan Jenis Mangga Berdasrkan Daun Dengan Metode K-Nearest Neighbor
Fakultas – Program Studi	: Fakultas Teknik- Teknik Informatika
Nama Perguruan Tinggi	: Universitas Nusantara PGRI Kediri
Alamat Perguruan Tinggi	: Jl. KH. Achmad Dahlan 76 Kediri

Dengan ini menyatakan bahwa :

a. artikel yang saya tulid merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;

b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui,		Kediri, Agustus 2017
Pembimbing I	Pembimbing II	Penulis,
Ardi Sanjaya, M.Kom NIDN : 0706118101	Julian Sahertian, S.Pd, M.T NIDN : 0707079001	Arif Agus Triono NPM: 13.1.03.02.0224

SISTEM PENGENALAN JENIS MANGGA BERDASARKAN DAUN METODE K-NEAREST NEIGHBOR

Arif Agus Triono
13.1.03.02.0224
Fakultas teknik-Teknik Informatika
arifagustriono93@gmail.com
Ardi Sanjaya, M.Kom , Julian Sahertian, S.Pd, M.T

UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Penelitian ini di latarbelakangi dari pengamatan dan pengalaman peneliti, bahwa daun mangga memiliki bentuk yang tidak sama meskipun memiliki sifat yang identik. Sehingga sering terjadi kesulitan dalam memberikan penilaian validitas daun mangga untuk menentukan jenis mangga.

Permasalahan penelitian ini adalah (1) Bagaimana merancang dan membangun sistem yang dapat digunakan sebagai pengenalan jenis mangga berdasarkan daun?. (2) Bagaimana menerapkan metode *K-Nearest Neighbor* untuk melakukan ekstraksi fitur pada citra digital?. (3) Bagaimana menerapkan metode *Euclidean Distance* untuk melakukan pencocokan pola pada citra digital?.

Penelitian ini menggunakan metode *K-Nearest Neighbor* sebagai pengekstrasi fitur citra, serta metode *Euclidean Distance* sebagai perhitungan jarak. Sebelum dilakukan pengenalan, *file* citra sebelumnya dilakukan proses *preprocessing* terlebih dahulu. Proses tersebut meliputi *grayscalling*, binerisasi, serta pendeteksian tepi *laplacian*.

Dari hasil pengujian pada skenario pertama, diperoleh akurasi sebesar 87% dengan menggunakan 80 data *training* dan uji sebanyak 5 kali *testing* dengan 16 daun mangga yang berbeda. Pada pengujian skenario kedua, diperoleh akurasi sebesar 81%. Pada skenario ketiga diperoleh akurasi sebesar 69% dengan menggunakan 48 data *training* dan uji sebanyak 5 kali *testing* dengan 16 daun mangga yang berbeda. Pengujian pada skenario keempat, diperoleh akurasi sebesar 62% dengan menggunakan 36 data *training* dan uji sebanyak 5 kali *testing* dengan 16 daun mangga yang berbeda. Uji testing ke lima di dapat nilai akurasi 55%. Dari hasil pengujian yang dilakukan, dapat disimpulkan bahwa metode *K-Nearest Neighbor* dapat digunakan untuk mengidentifikasi jenis mangga berdasarkan daun pada citra digital.

KATA KUNCI : daun manga, *laplace*, *K-Nearest Neighbor*

I. LATAR BELAKANG

Terdapat beranekaragam jenis tanaman yang tumbuh dan berkembang di permukaan bumi ini, sehingga masih banyak orang yang tidak mengetahui nama-nama dari tanaman tersebut, dan kemudian menimbulkan gagasan untuk menciptakan suatu sistem guna mengenal dan mempelajarinya. Beberapa ahli Biologi mencoba menciptakan suatu sistem untuk memper mudah mengenal dan mempelajari organisme melalui suatu cara pembuatan klasifikasi.

Pengklasifikasian merupakan proses pengelompokan berdasarkan ciri-ciri tertentu. Oleh karena itu, diperlukan aplikasi untuk pengelompokan daun jenis tanaman untuk membantu orang dalam mengenali dan mempelajari tentang tanaman.

Dalam kehidupan sehari-hari, manusia sering melihat tanaman di sekitarnya dengan ciri-ciri yang beranekaragam.

Keaneka ragaman tanaman tampak pada daunnya. Oval, waru (*cordate*), elips adalah beberapa bentuk dari daun. Akan tetapi, pada kenyataan masih banyak orang masih belum mengenal nama-nama tanaman tersebut karena jenis tanaman yang berane karagam. Seiring dengan berkembangnya kemajuan dibidang teknologi informasi, maka di mungkinkan untuk mengembangkan aplikasi guna melakukan pengenalan daun mangga yang di kelompokan dan dapat membantu orang dalam mengenali jenis daun tersebut.

Berbagai peneliti dalam pengenalan pola daun telah banyak di lakukan oleh para peneliti. Dan telah banyak juga metodologi yang telah dibangun untuk pengenalan pola daun. Febri Liantoni (2015) telah meneliti dan didalam penelitianya menjelaskan tentang penelitianya dengan judul "Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor" untuk mencari nilai kedekatan, didalam penelitian tersebut didapat tingkat akurasi tertinggi 86,67% tingkat akurasi dipengaruhi oleh nilai *threshold* (nilai ambang batas keabuan) dan juga perbandingan jumlah data *traning* dan data *testing* yang di gunakan. Pada tahun 2015 Sebastian Rori Listyanto menggunakan metode *K-Nearest Neighbor* untuk mengenali pola citra dalam mendeteksi penyakit kulit dengan tingkat keberhasilan tertinggi sebesar 85%.

Oleh karena itu dari beberapa uraian di atas penulis tertarik untuk merancang sebuah sistem pengenalan pola citra dengan menggunakan metode *K-Nearest Neighbor* (KNN) dalam sistem pengenalan jenis mangga bedasarkan daunnya.

II. Metode

A. Citra Digital

Menurut Sitorus, dkk (2006) Citra digital dibentuk dari sejumlah elemen terbatas, yang masing-masing elemen tersebut memiliki nilai dan koordinat tertentu. Menurut Darma, citra digital adalah citra *f(x,y)* yang telah di lakukan digitalisasi baik koordinat area maupun *brightness* level. Citra digital dipresentasikan sebagai matriks berukuran N x M dimana N= barisdan M= kolom. Masing-masing elemen pada citra digital (berarti elemen matriks) disebut *image element* atau *piksel*. Beberapa elemen yang penting dalam citra digital yaitu, kecerahan *(brightness),* kontras *(contrast),* kontur *(contour),* warna *(color),* bentuk *(shape),* dan tekstur *(texture).*

B. Citra *Grayscale*

Menurut Ahmad (2015) Citra *grayscale* memiliki pengertian sebagai berikut:

Citra warna *grayscale* menggunakan warna tingkatan warna abu-abu. Warna abu-abu merupakan satu-satunya warna pada ruang *RGB* dengan komponen merah, hijau, dan biru mempunyai intensitas yang sama. Banyaknya warna yang ada tergantung pada jumlah *bit* yang disediakan di memori untuk menampung kebutuhan warna ini.

Grayscaling citra merupakan tahapan pertama dari proses penyelarasan, pada tahap ini terjadi pengkonversian citra warna RGB menjadi citra berwarna abu-abu. Citra warna RGB terdiri dari 3 parameter warna yaitu merah *(red)*, hijau *(green)* dan biru *(blue)*, jika citra warna RGB ini dimasukan ke dalam proses ekstraksi, maka proses tersebut akan sulit untuk dilakukan karena citra *RGB* terdiri dari 3 parameter, oleh karena itu diperlukan penyamaan parameter yaitu dengan melakukan tahap *grayscaling* ini. (Julian, 2017)

Citra *grayscale* disimpan dalam format 8 bit untuk setiap sample pixel, yang memungkinkan sebanyak 256 intensitas. Format ini sangat membantu dalam pemrograman karena manipulasi bit yang tidak terlalu banyak. Untuk mengubah citra berwarna yang mempunyai nilai matrik masing-masing R, G dan B menjadi citra *grayscale* dengan nilai X, maka konversi dapat dilakukan dengan mengambil rata-rata dari nilai R, G dan B.

Berikut ini adalah persamaan untuk menghitung citra *grayscale*:

$X = \frac{R+G+B}{3}$ .....(1)

X = Nilai intensitas citra *grayscale*

R = Nilai intensitas warna merah dan citra asal 2

G = Nilai intensitas warna hijau dan citra asal

B = Nilai intensitas warna biru dan citra asal

C. *K-Nearest Neighbor* (KNN)

Klasifikasi dengan K-NN adalah mengelompokkan data baru, dalam penelitian ini adalah data uji berdasarkan jarak ke beberapa data ketetangga terdekat terhadap datalatih. Nilai k yang digunakan dalam penelitian adalah k= 1 dan k = 3. Prinsip kerja KNN adalah menghitung jarak menggunakan jarak *Euclidean,* jarak *Euclidean* digunakan untuk menghitung jarak antara dua vector yang berfungsi meng uji ukuran yang biasa digunakan sebagai interpretasi kedekatan jarak antara dua obyek yang direpresentasikan dalam persamaan. Dengan d(**x,y**):jarak *Euclidean* antara v ektor ***x*** dan vector ***y*** **;** x_i: fitur ke *i* dari vector.

$d(x,y) = \sqrt{\sum_i^n = 1(xi - yi)^2}$ .....(2)

x **;** y_i :fitur ke *i* dari vector **y ;** n: jumlah fitur pada vektor ***x*** dan ***y*** .

D. *Euclidean distance*

Menurut Santosa (2007). Jarak menjadi aspek penting dalam pengembangan metode pengklasifikasian maupun regresi. *Euclidean Distance* (jarak *Euclidean*) adalah jarak terpendek antara dua buah titik. Jika terdapat dua buah titik, maka jarak terpendek tersebut didapatkan dengan cara menarik garis lurus yang menghubungkan kedua titik tersebut. Dalam ruang *Euclidean* berdimensi n, R^n, jarak antara titik x dan y dapat dirumuskan berupa:

$$D = \sqrt{\sum_{i=1}^{n} | x_i - y_i |^2} \text{ .....(3)}$$

Keterangan :

D = Jarak *Euclidean*

n = Jumlah titik dalam R^n

Bila bilangan bekerja dalam ruang *Euclidean* berdimensi dua, maka jarak *Euclidean* antara titik $p(x_1, y_1)$dan $q(x_2, y_2)$ dapat dihitung dengan menggunakan rumus berikut :

$$D_{(p,q)} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \text{.... (4)}$$

Keterangan :

D = Jarak *Euclidean*

x_1 = Nilai *x* pada titik p

y_1 = Nilai *y* pada titik p

x_2 = Nilai *x* pada titik q

y_2 = Nilai *y* pada titik q

Keunggulan dari penggunaan metode *Euclidean Distaance* menurut Santosa (2007) dalam melakukan perhitungan pencocokkan antara lain Mudah dipahami dan diimplementasikan. Metode berjalan cukup dengan mendefinisikan fungsi untuk menghitung jarak antar-*instance*, menghitung jarak x dengan semua *instance* lainnya berdasarkan fungsi tersebut, dan menentukan nilai x sebagai nilai yang paling kecil.

a. Sangat *non-linear*. Garis keputusan kelas yang dihasilkan model perhitungan metode ini bersifat sangat fleksibel dan *nonlinear*.
b. Nilai kebenaran yang tinggi.Apabila kondisi dalam *dataset* tidak terdapat *noise*, prosentase hasil identifikasi dapat mencapai 100% meski dalam jumlah data yang banyak.

Sedangkan kekurangan dari penggunaan metode *Euclidean Distance* dalam melakukan perhitungan pencocokkan antara lain :

a. Lambat saat melakukan prediksi.Untuk mengklasifikasi sebuah *instance* x, *Euclidean Distance* harus menghitung jarak antara x dengan semua *instance* lain dalam *dataset* yang ada. Dengan kata lain, kompleksitas waktu klasifikasi *Euclidean Distance* berbanding lurus dengan jumlah *instance* latih. Jika *dataset* yang ada berukuran besar (terdiri dari banyak *instance* dan/atau banyak 6ariable), prosesnya 6ari jadi sangat lambat.

b. Rentan terhadap perbedaan rentang 7ariable. Dalam perhitungan jarak antar-*instance*, *Euclidean Distance* menganggap semua 7ariable setara atau sama penting(bagian penjumlahan pada rumus perhitungan jarak). Jika terdapat satu variabel yang memiliki rentang jauh lebih besar dibanding variabel-variabel lainnya, maka perhitungan jarak akan didominasi oleh *variabel tersebut*.
c. Rentan terhadap dimensionalitas yang tinggi. Hal ini karena semakin banyak dimensi, ruang yang bisa ditempati *instance* semakin besar, sehingga semakin besar pula kemungkinan bahwa jarak kedekatan dari suatu *instance* sebetulnya sama sekali tidak memiliki kedekatan.
d. Tidak menangani nilai hilang (*missing value*) secara implisit. Jika terdapat nilai hilang pada satu atau lebih variabel dari suatu *instance*, perhitungan jarak *instance* tersebut dengan *instance* lainnya menjadi tidak terdefinisi.

E. Deteksi laplacian

Menurut Madenda (2015:110) Operator *Laplace* biasanya selalu disebut dengan sebutan operator turunan. Operator *Laplace* biasanya mendeteksi tepi lebih akurat khususnya pada tepi yang curam. Pada tepi yang curam turunan keduanya memiliki *Zercrossing* (persilangan nol), yaitu titik dimana terdapat pergantian tanda nilai turunan kedua, sedangkan pada tepi yang ditandai tidak terdapat persilangan nol.

1. Representase turunan kedua dalam bentuk.
2. Kernel operator Laplacian adalah sebagai berikut :

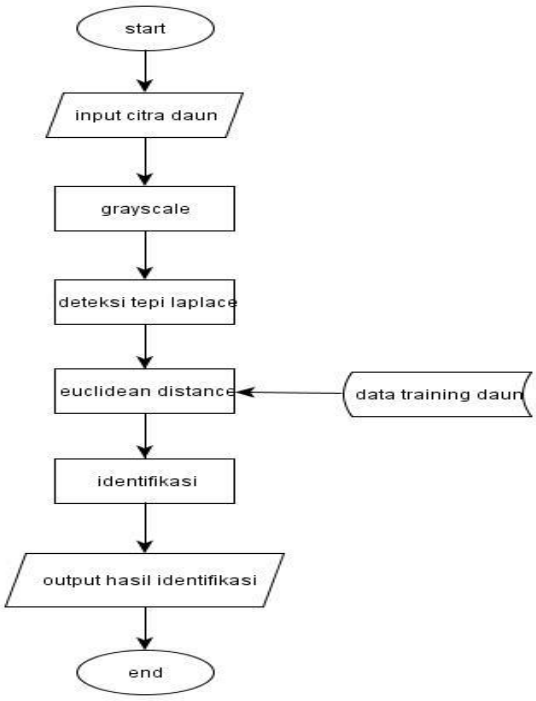
L=

0	-1	0
-1	4	-1
0	-1	0

III. HASIL DAN KESIMPULAN

A. Perancangan Sistem

Berikut *Flowchart* sistem berfungsi memberikan gambaran tentang alur program yang dibuat. Pada bagian ini akan dijelaskan bagaimana proses pengolahan data yang berupa citra, kemudian diolah menggunkan pengolahan citra sehingga dapat dihasilkan kemampuan mengidentifikasi suatu objek. Berikut merupakan gambaran *flowchart* sistem.



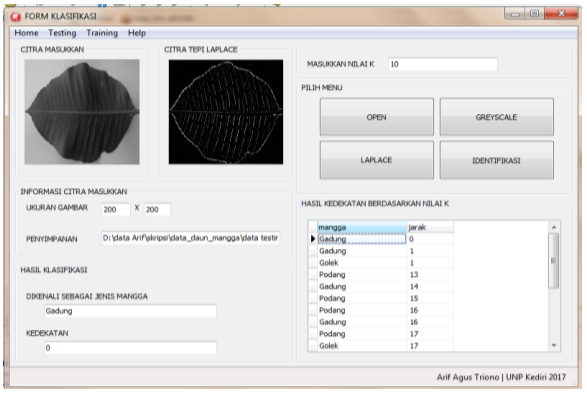
Gambar 1. *Flowchart* sistem

a. *Star* : Aplikasi dijalankan.
b. *Input* : Masukkan citra daun mangga yang akan di indentifikasi.
c. *Grayscale* : Merubah citra daun menjadi *grayscale*
d. Deteksi Tepi *Laplace* : Melakukan proses deteksi tepi *laplace*
e. *Euclidean Distance* : Menghitung nilia jarak dari yang diinputkan oleh *user* dengan semua data citra daun mangga yang terdapat pada data *training*, kemudian nilai jarak tersebut disimpan.
f. Identifikasi : Mencari nilai jarak terkecil yang selanjutnya akan dijadikan sebagai acuan sebagai hasil identifikasi.
g. *Output* : *Output* yang dihasilkan adalah hasil seteksi tepi *laplace* dan hasil identifikasi *euclidean distance*
h. *End* : Aplikasi selesai dijalankan.

B. Tampilan Program

Pada sistem pengenalan jenis mangga yang dibangun,tampilan program dibuat dengan *desain* yang sederhana dengan tujuan untuk mempermudah penggunaannya. Berikut tampilan program yang dibuat:

1. Form Data Testing



Gambar 2. Tampilan *Form Testing*

Keterangan dari gambar 2 diatas adalah sebagai berikut :

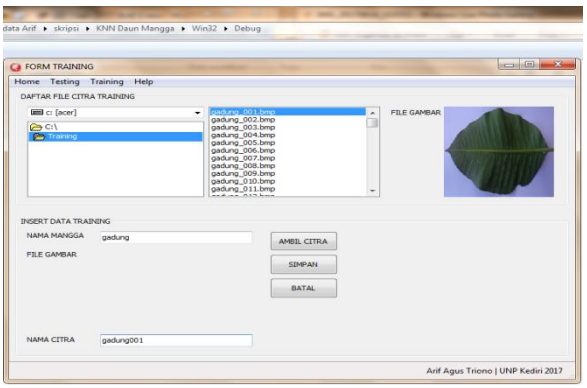
Pada *form testing*, *user* dapat memilih citra yang akan di identifikasi dengan cara menekan *open* untuk memilih citra testing yang sudah di simpan di dalam folder, dalam gambar 5.6 yang dipilih adalah citra dengan nama gadung001.bmp yang terdapat di *drive* d dengan nama folder data daun *testing*, tampilan dari citra tersebut akan tampil pada *box* citra RGB.

Selanjutnya pada citra tersebut dilakukan proses *grayscale* dan deteksi tepi *laplace*, hasil dari deteksi tepi akan tampil pada *box* hasil identifikasi *laplace*. Tombol

identifikasi berfungsi untuk memulai proses identifikasi, kemudian setelah proses identifikasi pada *box* akan muncul nama jenis mangga gadung.

2. *Form* Data *Training*

Form data *training* yaitu *form* yang di gunakan *user* untuk memasukkan gambar data *training* yang baru ke dalam *database*. Berikut tampilan dalam data *training :*



Gambar 3 Tampilan *Form Training*

Keterangan dari gambar 3 adalah sebagai berikut :

Pada *form training*, *user* dapat melihat isi dari *database* gambar (data *training*) dalam program ini *database* gambar terletak pada C:\citra training, kemudian pilih citra dan akan tampil pada *box* citra. Selanjutnya pada citra tersebut dilakukan penyimpana ke dalam *database*, kemudian *box* batal untuk proses batal dalam penyimpanan.

C. Hasil Uji Coba Sistem

Dalam pengujian sistem pengenalan daun mangga ini dilakukan 5 kali uji coba, dan disetiap uji coba memiliki jumlah data *training* yang berbeda, Untuk memperoleh nilai akurasi pada setiap uji coba menggunakan rumus :

$$\frac{\text{jml hasil benar}}{\text{jml data testing}} \times 100 \quad \text{.....(5)}$$

Berikut merupakan tabel hasil uji coba dari 5 skenario pengujian pengenalan jenis mangga

Tabel Uji coba 1

Skenario	Jumlah Citra		Hasil Identifikasi		Akurasi
	Training	Testing	Benar	Salah	
1	80	16	15	2	87 %
2	64	16	13	3	81 %
3	48	16	11	5	69 %
4	36	16	10	6	62 %
5	16	16	9	7	55 %

Keterangan dari tabel 5.1. adalah sebagai berikut :

a. Pada skenario 1, jumlah data *training* 80 dan data *testing* 16. Hasil identifikasi 15 benar, dan 2 salah. Maka nilai akurasi 87 %.
b. Pada skenario 2, jumlah data *training* 64 dan data *testing* 16. Hasil identifikasi 13 benar, dan 3 salah. Maka nilai akurasi 81 %.
c. Pada skenario 3, jumlah data *training* 48 dan data *testing* 16. Hasil identifikasi benar 11, dan salah 5. Maka nilai akurasi 69 %.
d. Pada skenario 4, jumlah data *training* 36 dan data *testing* 16. Hasil identifikasi 10 benar , dan 6 salah. Maka nilai akurasi 62 %.
e. Pada skenario 5, jumlah data *training* 16 dan data *testing* 16. Hasil identifikasi

benar 9, dan salah 7. Maka nilai akurasi 55 %.

Tabel Uji Coba 2

No	Daun Mangga	Hasil Uji Coba				
		I	II	III	IV	V
1	Gadung	Gadung	Gadung	Gadung	Golek	Gadung
2	Golek	Golek	Golek	Golek	Golek	Gadung
3	Manalagi	Manalagi	Manalagi	Podang	Podang	Podang
4	Podang	Podang	Podang	Podang	Manalagi	Manalagi
5	Gadung	Golek	Gadung	Gadung	Golek	Golek
6	Golek	Golek	Golek	Gadung	Golek	Golek
7	Manalagi	Manalagi	Podang	Manalagi	Manalagi	Podang
8	Podang	Podang	Podang	Manalagi	Podang	Podang
9	Gadung	Gadung	Golek	Gadung	Gadung	Gadung
10	Golek	Golek	Golek	Golek	Gadung	Gadung
11	Manalagi	Manalagi	Manalagi	Golek	Manalagi	Manalagi
12	Podang	Manalagi	Podang	Podang	Podang	Podang
13	Gadung	Gadung	Gadung	Gadung	Gadung	Gadung
14	Golek	Golek	Golek	Golek	Golek	Golek
15	Manalagi	Manalagi	Manalagi	Manalagi	Manalagi	Manalagi
16	Podang	Podang	Podang	Podang	Podang	Podang
	Akurasi	87%	81%	69%	62%	55%

Kesimpulan Dari ke lima kali data *testing* dengan daun yang berbeda keterkaitanya . Dan data pertama merupakan hasil nilai akurasi tertinggiyang didapatkan setelah lima kali uji coba dilakukan dengan jumlah data *training* yang berbeda. Adapun hasil akurasi program yang dihasilkan di pengaruhi oleh jumlah data *training* yang inputkan. Selain itu data *testing* yang digunakan pada penelitian ini merupakan hasil dari pemotretan yang mana dari segi bentuk maupun ketepatan masing-masing menjadi faktor penentu tinggi rendahnya akurasi program itu sendiri.

D. Kesimpulan

Dari hasil analis pada bab sebelumnya dapat diambil kesimpulan sebagai berikut:

1. Aplikasi ini sudah mampu berjalan sesuai dengan perancangan sistem dan mampu mengelali jenis daun mangga.
2. Metode *K-Nearest Neighbo*r bisa di implementasikan untuk sistem pengenalan jenis mangga berdasarkan daun dan hasil uji coba pengenalan yang di dapat bisa mencapai akurasi 87%. Akurasi hasil pencocokan juga di pengaruhi oleh pemotretan itu sendiri. Baik itu dari segi sisi posisi pengambilan gambar karena mengingat dalam daun memiliki tulang yang hampir sama.
3. Performa hasil kerja dari aplikasi pengenalan jenis mangga berdasarkan daun sudah cikup baik dan berjalan lancar.

IV. DAFTAR PUSTAKA

Darma, Putra. 2010. *Pengolahan Citra Digital(Westriningsih,Ed.)*.Yogy karta: Penerbit ANDI.

Elektronika Dasar. 2012. *Definisi dan Pengolahan Citra Digital*, dalam http://www.elektronika-dasar.web.id. Diakses Pada 25 Mei 2017.

Fadlisyah, Fauzan, Taufiq, Zulfikar. 2008. *Pengolahan Citra Menggunakan Delphi*. Yogyakarta: Graha Ilmu.

Sahertian, Julian, and Sanjaya, Ardi. "*Deteksi Buah pada Pohon Menggunakan Metode SVM dan Fitur Tekstur*." Semnasteknomedia. Online 5.1 (2017): 4-3

Kadir, Abdul. 2013. Dasar Pengolahan Citra dengan Delphi (Dewiberta Hardjono, Ed.). Yogyakarta: Penerbit ANDI.

Putra, Darma. 2009. *Euclidean Distance.* Yogyakarta: Graha ilmu.

Portal Pertanian. 2014. Macam-macam Buah Mangga,*(online)*.Dalam http://kebunq.com/2014/09/macam-macam-buah-mangga.html,Diakses pada 25 Mei 2017.

Rori Sebastian, Listyanto. 2015. Implementasi *K-Nearest Neighbor* untuk mengenali pola citra dalam mendeteksi penyakit kulit. (Online).

Sufriadi,2010*Teori Dasar Pengolahan Citra* http://www.sufriadi.files.wordpress.com Diakses Pada 25 Nov 2016 14:00

Sitorus, Syahriol, dkk. 2006. Pengolahan Citra Digital. Universitas Sumatra Utara.