

# Optimasi Algoritma K-Nearest Neighbor untuk Klasifikasi Tingkat Kematangan Buah Alpukat Berdasarkan Warna

M. Habib Hanafi<sup>1</sup>, Nurul Fadillah<sup>2</sup>, Ahmad Ihsan<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Samudra<sup>1,2,3</sup>

Muhammadhabib@gmail.com<sup>1</sup>, nurulfadillah@unsam.ac.id<sup>2</sup>, ahmadihsan@unsam.ac.id<sup>3</sup>

---

## Article Info

### History :

Dikirim 27 Desember 2018

Direvisi 23 Februari 2019

Diterima 12 April 2019

---

### Kata Kunci :

Alpukat  
K-Nearest Neighbor  
Matang  
Mentah  
Setengah Matang

---

## Abstrak

Sistem komputerisasi semakin dibutuhkan seiring dengan perkembangan teknologi saat ini. Salah satu sistem komputerisasi yang butuh pengujian serta pengembangan sistem lebih lanjut ialah klasifikasi citra digital. Salah satu topik yang dapat diangkat ialah tentang tingkat kematangan buah. Alpukat (*Persea americana* mill) merupakan tanaman yang dapat tumbuh subur di daerah tropis seperti Indonesia dan merupakan salah satu jenis buah yang digemari masyarakat. Alpukat Memiliki nilai warna yang mirip disetiap tingkat kematangan buah, menjadi hal yang menarik untuk dibahas dan menjadi fokus utama dalam penelitian ini. Penelitian ini mendeskripsikan optimasi dengan menggunakan algoritma K-Nearest Neighbor Classifier untuk memperoleh hasil klasifikasi yang lebih efektif. Tujuan dari penelitian ini ialah menguji keakuratan dari optimasi algoritma k-nearest neighbor pada klasifikasi tingkat kematangan buah alpukat. Pengujian dilakukan dengan memperoleh nilai data training dan melakukan uji sistem. Data uji training menggunakan tiga objek jenis alpukat yang berbeda yaitu matang, setengah matang dan mentah. Sedangkan pengujian pada sistem dilakukan dengan melakukan input data berupa 14 sampel alpukat ke sistem yang telah dirancang. Sample buah alpukat berupa gambar berekstensi .jpg dan .png yang terdiri dari 4 sampel alpukat setengah matang. 5 sampel alpukat matang dan 5 sampel alpukat mentah. Hasil dari optimasi dari algoritma ini memperoleh nilai keakuratan sebesar 78,56 %. Selain itu, optimasi algoritma ini mampu meminimalisir hasil kesalahan prediksi akibat nilai range diluar jangkauan dan mampu mengoptimalisasi bobot Persentase Nilai Training setiap uji citra dalam menentukan hasil klasifikasi.

© This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

---

## Koresponden:

M. Habib Hanafi,

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Samudra

Pasar-III Datuk Kabu, Dusun XV GG. Melati Raya No.11, Deli Serdang , Indonesia, 20371

Email : Muhammadhabib348@gmail.com

---

## 1. PENDAHULUAN

Klasifikasi citra merupakan suatu proses pengelompokan seluruh *pixel* pada suatu citra kedalam dalam kelompok sehingga dapat diinterpretasikan sebagai suatu *property* yang spesifik

[1]. Kehadiran klasifikasi citra digital telah menghasilkan berbagai penelitian yang menggunakan beragam algoritma. Namun, penelitian mengenai optimasi algoritma, khususnya pada pengujian objek yang memiliki nilai uji yang hampir sama disetiap objeknya belum ditemukan secara spesifik saat ini. Optimasi algoritma tentu diperlukan, khususnya pada objek data uji yang memiliki nilai yang hampir sama. Objek pengujiannya pun seharusnya menggunakan objek yang memiliki tingkat kemiripan yang hampir sama, baik dalam hal bentuknya ataupun dari warnanya.

Alpukat (*Persea americana* mill) merupakan tanaman yang dapat tumbuh subur di daerah tropis seperti Indonesia dan merupakan salah satu jenis buah yang digemari masyarakat karena selain rasanya yang enak juga kandungan antioksidannya yang tinggi [2]. Buah alpukat merupakan buah yang sering kita jumpai. Buah serbaguna ini memiliki banyak manfaat dan khasiat bagi manusia. Ada banyak zat yang kaya manfaat yang terdapat di buah ini.. Rasanya yang nikmat membuat banyak orang menyukainya. Seperti buah pada umumnya, alpukat memiliki tingkat kematangan tersendiri. Kematangan buah alpukat dapat dilihat dari warna kulit maupun warna daging buahnya. Kemampuan akan membedakan tingkat kematangan buah alpukat tentu saja dibutuhkan pengamatan yang akurat. Hal ini disebabkan karena alpukat memiliki warna yang hampir mirip antara alpukat matang dengan setengah matang.

Klasifikasi tingkat kematangan alpukat bisa diwujudkan melalui perancangan algoritma yang menyesuaikan parameter dari objek alpukat tersebut. Jenis kematangan alpukat sendiri bisa dibedakan berdasarkan warna dan bentuknya. Atas dasar itulah penulis melakukan penelitian tentang optimasi algoritma *K-Nearest Neighbor* dengan objek buah alpukat untuk menghasilkan klasifikasi kematangan buah berdasarkan warna . Metode *K-Nearest Neighbor* penulis pilih karena memiliki kelebihan yang diantaranya ialah metode ini KNN memiliki beberapa kelebihan yang diantaranya ialah algoritma ini tangguh terhadap training data yang noisy dan efektif apabila data latih nya besar. Selain mengembangkan klasifikasi citra digital yang mampu membedakan jenis kematangan alpukat, penelitian ini juga bertujuan untuk mengukur sejauh mana keakuratan dari optimasi yang telah ditambah pada metode *K-Nearest Neighbor* untuk klasifikasi alpukat berdasarkan warnanya.

## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Pengumpulan data merupakan tahap yang bertujuan untuk memperoleh informasi- informasi atau data-data yang berhubungan dengan penelitian. Pada pengumpulan data digunakan dua metode pengumpulan data yaitu studi pustaka dan observasi. Studi pustaka dilakukan untuk menemukan dan mengumpulkan data dan informasi tentang alpukat, metode *k-nearest neighbour* serta mendapatkan perolehan data berupa teknik pengolahan citra digital. Setelah studi pustaka, langkah selanjutnya ialah melakukan observasi. Observasi dilakukan ke pasar-pasar di kota langsa untuk mengumpulkan data buah alpukat. Setelah mendapatkan data tersebut, buah alpukat tersebut di foto dengan memiliki latar background putih.

### 2.2 Konsep Teori

#### 2.2.1 K-Nearest Neighbor

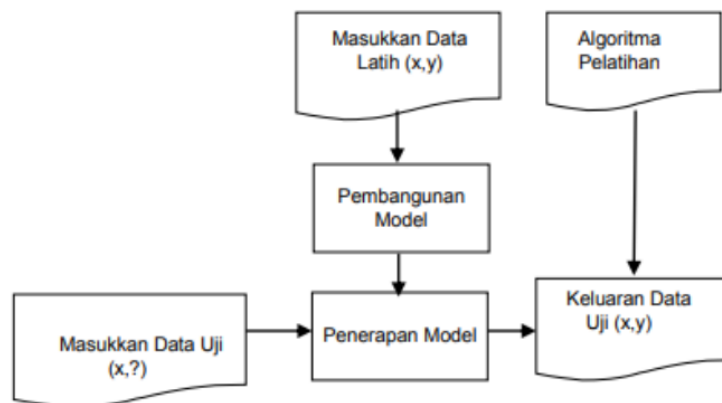
Algoritma K-Nearest Neighbor merupakan metode klasifikasi yang mengelompokan data baru berdasarkan jarak data baru itu kebeberapa data/tetangga (*neighbor*) terdekat [3]. K-Nearest Neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada [4].

#### 2.2.2 Konsep Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang ilakukan, yaitu : pertama, Pembangunan model sebagai *prototype* untuk disimpan sebagai memori dan kedua, Penggunaan model tersebut untuk melakukan pengenalan/ klasifikasi/ prediksi pada

suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang mudah disimpan [5].

Contoh aplikasi yang sering ditemui adalah pengklasifikasian jenis hewan, yang mempunyai sejumlah atribut. Dengan atribut tersebut, jika ada hewan baru, kelas hewannya bisa langsung diketahui. Contoh lain adalah bagaimana melakukan diagnosis penyakit kulit kanker melanoma, yaitu dengan melakukan pembangunan model berdasarkan data latih yang ada, kemudian menggunakan model tersebut untuk mengidentifikasi penyakit pasien baru sehingga diketahui apakah pasien tersebut menderita kanker atau tidak. Klasifikasi pada umumnya menerapkan proses kerja sebagai berikut :



Gambar 1. Proses Klasifikasi

### 2.2.3 Pengukuran Kinerja Klasifikasi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi. Kuantitas matriks konfusi dapat diringkus menjadi dua nilai, yaitu akurasi dan laju error. Dengan mengetahui jumlah data yang diklasifikasikan secara benar, kita dapat mengetahui akurasi hasil prediksi dan dengan mengetahui jumlah data yang diklasifikasikan secara salah, kita dapat mengetahui laju error dari prediksi yang dilakukan. Dua kuantitas ini digunakan sebagai matrik kinerja klasifikasi. Untuk menghitung akurasi digunakan formula.

### 2.2.4 K-Nearest Neighbor Classifier

*K-Nearest Neighbor Classifier* (k-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang dimensi dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Nilai k yang terbaik untuk algoritma ini tergantung pada data, secara umum nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, akan tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai k yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*. Kasus khusus di mana klasifikasi diprediksikan berdasarkan data pembelajaran yang paling dekat (dengan kata lain,  $k = 1$ ) yang biasanya disebut algoritma nearest neighbor [6]. Rumus penghitungan jarak ditulis sebagai berikut:

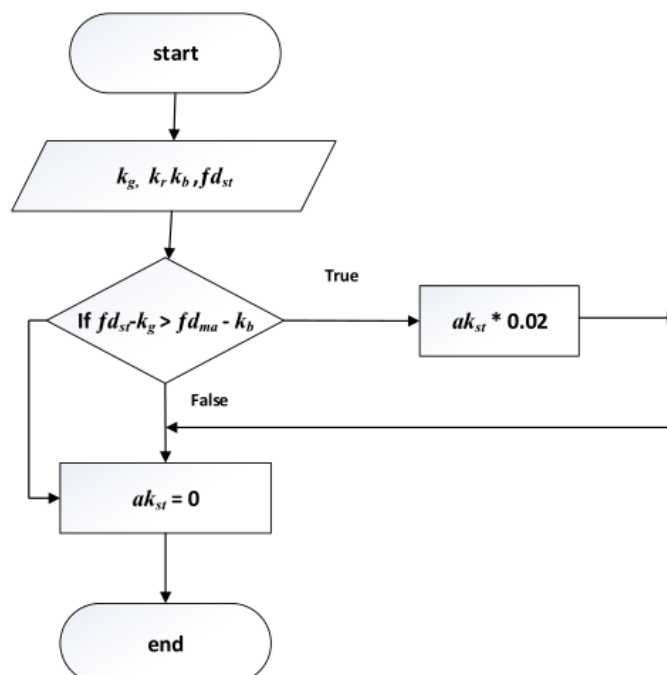
## 2.2.5 Ekstraksi Ciri

Setelah memperoleh data buah alpukat berupa gambar berformat .jpg, tahap selanjutnya ialah melakukan ekstraksi ciri. Proses ekstraksi ciri bertujuan untuk mengambil ciri pada citra agar dapat diproses ke dalam proses klasifikasi. Ekstraksi ciri pada penelitian ini ialah merupakan proses perolehan nilai *Red*, *Green* dan *Blue*.

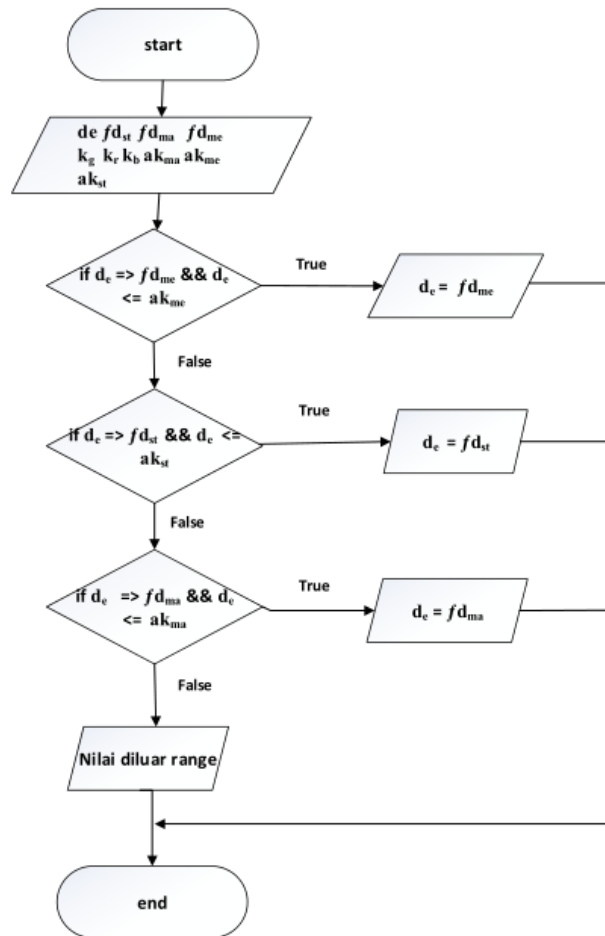
## 2.2.6 Optimasi Algoritma K-Nearest Neighbor

Optimasi adalah upaya atau cara untuk memperoleh hasil yang terbaik [7]. Adapun optimasi algoritma K-Nearest Neighbor pada penelitian ini dilakukan dengan menentukan nilai range serta nilai alternatifnya pada nilai  $fd_i$  yang merupakan nilai perolehan dari data buah training. Langkah optimasi ini dilakukan karena pada umumnya perolehan nilai  $fd_i$  merupakan nilai yang tidak pasti sehingga akan menyebabkan hasil perolehan klasifikasi yang kurang optimal. Langkah dilakukan dengan menoperasikan berbagai variable seperti : setengah matang ( $fd_{st}$ ) training matang ( $fd_{ma}$ ), training mentah ( $fd_{me}$ ) nilai green ( $k_g$ ) nilai red ( $k_r$ ) nilai blue ( $k_b$ ) alternatif nilai matang ( $ak_{ma}$ ) alternatif nilai mentah ( $ak_{me}$ ) alternatif nilai setengah matang ( $ak_{st}$ ). Adapun proses pada optimasi algoritma ini ialah sebagai berikut.

1. Menentukan nilai alternatif. Kondisi nilai citra yang hampir mirip dari setiap nilai training objek menyebabkan ketiga nilai training dari Red, Green, Blue jika dibagi tentu akan menghasilkan rata-rata yang belum tentu sesuai, padahal buah alpukat masih dikatakan setengah matang jika masih terdapat corak warna hijau dipermukaan buahnya meskipun nilainya sangat sedikit. Sehingga menyebabkan harus adanya toleransi nilai tambahan. Proses pada optimasi langkah ini yaitu dengan menentukan apakah  $fd_{st} \cdot k_g > fd_{ma} - k_b$ . Jika kondisi true, maka akan memproses  $ak_{st} * 0.02$ . Jika kondisi false, maka  $ak_{st} = 0$ . Adapun proses langkah optimasi ini dijelaskan melalui flowchart pada gambar 2.
2. Langkah kedua dalam proses optimasi ialah menentukan nilai range untuk menghasilkan hasil klasifikasi yang sesuai dengan jangkauan nilai training sehingga dapat mengoptimalkan hasil klasifikasi yang lebih akurat dan mengantisipasi perolehan hasil klasifikasi yang salah prediksi. Hasil dari optimasi langkah kedua ini akan memberikan hasil akhir dari proses klasifikasi. Adapun proses langkah optimasi ini dijelaskan melalui flowchart pada gambar 3.



Gambar 2. Proses Optimasi Langkah Kedua



Gambar 3. Proses Optimasi Langkah Ketiga

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengujian Data Training

Perolehan data training di peroleh dari ekstraksi citra mejadi nilai RGB. Hasil dari RGB ini akan menjadi acuan data untuk masing-masing jenis alpukat. Adapun perolehan data training ialah sebagai berikut. Dari data diatas, perolehan K yang merupakan hasil RGB untuk masing jenis alpukat ialah 87000 untuk alpukat mentah, 30753 untuk alpukat matang dan 28457 untuk alpukat setengah matang. Adapun citra yang digunakan sebagai data uji training serta hasil pengujiannya ialah sebagai berikut..



(a) Alpukat Mentah



(b) Alpukat Setengah Matang



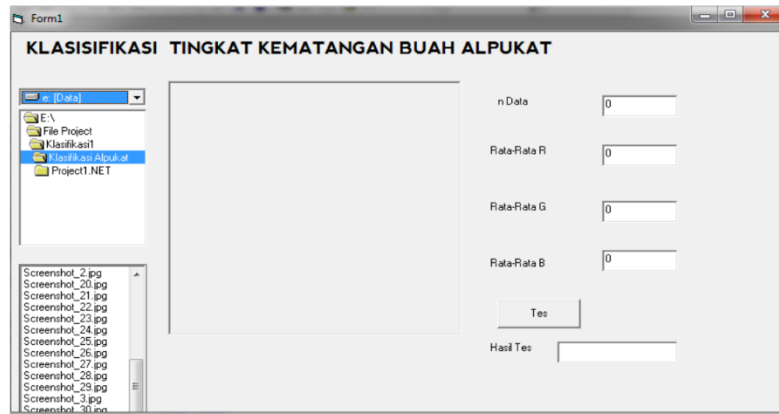
(c) Alpukat Matang

Nama Data	n Data	Rata-Rata R	Rata-Rata G	Rata-Rata B
<input type="radio"/> Mentah	87000	44.9384137931	64.0531609195	37.9635057471
<input checked="" type="radio"/> Matang	30753	39.6765518811	40.7697785581	45.0452638766
<input type="radio"/> Setengah Matang	28457	48.0677864848	51.1525108057	46.7828653758

Gambar 4. Perolehan Nilai Setiap Jenis Alpukat

### 3.2 Hasil Pengujian

Pengujian pada sistem dilakukan dengan melakukan input data berupa 14 sample alpukat ke sistem yang telah dirancang. Sample buah alpukat berupa gambar berekstensi .jpg dan .png yang terdiri dari 4 sampel alpukat setengah matang. 5 sampel alpukat matang dan 5 sampel alpukat mentah. Adapun gambar sistem serta hasil pengujian dari sistem ialah sebagai berikut.



Gambar 5. Tampilan Sistem Dari Uji Citra

Tabel 1. Pengujian Sampel Buah Alpukat

No.	Pengamatan Manual	Hasil Pengujian
1	Matang	K : 29000
		Jarak Nilai Training : 1753
		Persentase Nilai Training : 94%
		Hasil Sistem : Matang Keterangan : Akurat
2	Matang	K : 26884
		Jarak Nilai Training : 3869
		Persentase Nilai Training : 87%
		Hasil : Tidak Akurat
3	Matang	K : 30253
		Jarak Nilai Training : 500
		Persentase Nilai Training : 98%
		Hasil : Akurat
4	Matang	K : 28993
		Jarak Nilai Training : 1760
		Persentase Nilai Training : 93%
		Hasil Sistem : Matang Keterangan : Akurat

5	Matang	K : 29254
		Jarak Nilai Training : 1499
		Persentase Nilai Training : 95%
		Hasil Sistem : Matang
		Keterangan : Akurat
6	Setengah Matang	K : 28300
		Jarak Nilai Training : 157
		Persentase Nilai Training : 99%
		Hasil Sistem : Setengah Matang
		Hasil : Akurat
7	Setengah Matang	K : 19600
		Jarak Nilai Training : 8857
		Persentase Nilai Training : 68%
		Hasil Sistem : Nilai keluar dari range
		Hasil : Tidak Akurat
8	Setengah Matang	K : 21480
		Jarak Nilai Training : 6977
		Persentase Nilai Training : 75%
		Hasil Sistem : Nilai keluar dari range
		Hasil : Tidak Akurat
9	Setengah Matang	K : 26006
		Jarak Nilai Training : 2451
		Persentase Nilai Training : 91%
		Hasil Sistem : Setengah matang
		Hasil : Akurat
10	Mentah	K : 91383
		Jarak Nilai Training : 4383
		Persentase Nilai Training : 95%
		Hasil Sistem : Mentah
		Hasil : Akurat
11	Mentah	K : 87000
		Jarak Nilai Training : 0
		Persentase Nilai Training : 0%
		Hasil Sistem : Mentah
		Keterangan : Akurat
12	Mentah	K : 57000
		Jarak Nilai Training : 30000
		Persentase Nilai Training : 65%
		Hasil Sistem : Mentah
		Keterangan : Akurat
13	Mentah	K : 88020
		Jarak Nilai Training : 1020
		Persentase Nilai Training : 98%
		Hasil Sistem : Mentah
		Keterangan : Akurat
14	Mentah	K : 98626
		Jarak Nilai Training : 11626
		Persentase Nilai Training : 88%
		Hasil Sistem : Mentah
		Keterangan : Akurat

Berdasarkan hasil pengujian diatas, diperoleh tingkat akurasi sebesar 78,57% ( $11/14$ ) \*100% = 78,57%. Berdasarkan hasil pengujian tersebut, Data uji berupa citra matang dengan setengah matang mendominasi hasil ketidakakuratan. Hal ini dikarenakan nilai uji training terlalu berdekatan diantara keduanya. Untuk alpukat matang memiliki nilai training sebesar 30753 sedangkan untuk alpukat setengah matang nilai yang diperoleh sebesar 28457. Selisih nilai uji training tersebut hanya berkisar 2296 saja. Selain itu, ketidakakuratan disebabkan pembacaan nilai uji dari sistem yang mengkalkulasikan nilai klasifikasi yang sangat jauh dari nilai training. Meskipun algoritma k-nearest neighbor tangguh terhadap training data yang noisy dan efektif apabila data latih nya besar, algoritma ini belum teruji efektif pada kedua kondisi tersebut.

Meskipun demikian, hasil optimasi algoritma pada penelitian ini dapat membuat proses pengujian citra dalam meningkatkan hasil klasifikasi yang lebih efektif. Efektifitas tersebut dilihat dari hasil perolehan uji citra untuk nilai persentase nilai training 91% yang menghasilkan nilai akurat untuk alpukat setengah matang, padahal jarak nilai training antara dua objek antara alpukat matang dan setengah matang ialah 92% sehingga tanpa adanya optimasi ini hasil klasifikasi bias saja tidak akurat karena sistem akan menganggap bahwa nilai persentase nilai training 91% masih buah alpukat matang. Hal ini sangat baik untuk klasifikasi citra yang memiliki objek percobaan dengan ragam warna yang hampir mirip seperti alpukat matang dan setengah matang. Selain baik untuk citra yang memiliki warna harna hampir mirip, optimasi algoritma ini juga dapat menghasilkan kasifikasi yang lebih efektif untuk citra yang memiliki jarak nilai yang sangat jauh. Hal ini terbukti pada pengujian citra alpukat mentah yang dapat menghasilkan keakuratan meskipun persentase nilai training sebesar 65%. Hal yang juga membuat optimasi algoritma ini dikatakan efektif ialah memberikan keterangan saat nilai uji keluar dari range. Hasil ini tentu membuat sistem tidak sembarangan memberikan hasil klasifikasi meskipun pengujian tersebut belum menghasilkan klasifikai yang akurat.

#### 4. KESIMPULAN

Berdasarkan uraian pembahasan dari penelitian yang telah dilakukan maka dapat disimpulkan sebagai berikut:

1. Penerapan optimasi algoritma metode k-nearest neighbor untuk klasifikasi kematangan buah alpukat menghasilkan akurasi sebesar 78,57 %. Persentase tersebut diperoleh dari data sampel buah alpukat sebanyak 14 sampel.
2. Tingkat tidak keakuratan pada penelitian ini berasal dari sampel alpukat setengah matang dan matang. Hal ini dikarenakan warna dari data training pada sampel tersebut memiliki tingkat perolehan nilai RGB yang hampir sama.
3. Hasil optimasi mampu meminimalisir hasil kesalahan prediksi akibat nilai rengen diluar jangkauan dan mampu mengoptimalisasi bobot Persentase Nilai Training setiap uji citra dalam menentukan hasil klasifikasi.
4. Proses optimasi algoritma ini bukan hanya bisa diterapkan pada objek buah alpukat saja, melainkan dapat diterapkan objek yang lain seperti buah, hewan, benda dan lain-lain.
5. Dengan perolehan nilai keakuratan kurang dari 80 %, penulis menyarankan agar penelitian mengenai klasifikasi citra perlu ditingkatkan dengan mengembangkan algoritma yang mampu memberikan hasil klasifikasi yang lebih efektif untuk objek citra berupa warna-warna yang mirip disetiap objeknya Hal ini tentu akan menjadi bahan penelitian yang berguna bagi kepentingan manusia di masa depan.

#### DAFTAR PUSTAKA

- [1] Chein-I Chang and H.Ren.. “An Experiment-Based Quantitative and Comparative Analysis of Target Detection and Image Classification Algorithms for Hyperspectral Imagery”. IEEE Trans. on Geoscience and Remote Sensing. 2000.
- [2] Karina, Anna.” Khasiat dan Manfaat Alpukat”, edisi ke I. Surabaya: Setomata, 2012.



- 
- [3] Santoso, Singgih. "Statistik Deskriptif: Konsep dan Aplikasi dengan Microsoft Exel dan SPSS". Yogyakarta: ANDI, 2007.
  - [4] Kusriani, luthfi,E.T. "Algoritma Data Mining". Yogyakarta: ANDI, 2012.
  - [5] Prasety, Eko, "Data Mining Konsep dan Aplikasi Menggunakan MATLAB". Yogyakarta: ANDI, 2009.
  - [6] Liantoni and Nugroho. " Klasifikasi Daun Herbal Menggunakan Metode Naive Bayes Classifier dan K Nearest Neighbor " Jurnal Simantec Vol. 5, Des. 2015.
  - [7] KEMDIKBUD. 2019. Kamus Besar Bahasa Indonesia [Online]. Available : <https://kbbi.kemdikbud.go.id/entri/optimalisasi>