

KLASIFIKASI KUALITAS KESEGERAN BUAH SEMANGKA BERDASARKAN FITUR WARNA YCbCr MENGGUNAKAN ALGORITMA K-MEANS TERBOBOT

(*Watermelon Quality Classification Using Weighted K-means Algorithm Based on Features of YCbCr Color*)

Lalu Zulfikar Muslim*, I Gede Pasek Sutawijaya, Fitri Bimantoro

Dept Informatics Engineering, Mataram University

Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA

Email: zulvcarmoslem22@gmail.com, gpsutawijaya@unram.ac.id, bimo@unram.ac.id

Abstract

The classification of fruit quality on a computer using image data is very necessary for a faster and easier sorting process. Additionally, this can also be used in making decisions and policies related to business strategies in the industry. The paper presents the quality classification of watermelon that is carried out using the Weighted K-Means Algorithm. The watermelon is classified into three groups, namely fresh, medium, and rotten. There are two stages for the classification process, namely training, and examinations. The classification works using the YCbCr color space. In the training phase, the pair of input and the target of data that is processed to obtain the weight of k-Means. While for the testing/classification phase, the input data processed is an arbitrary image that has not been classified. The classification shows that the greater the amount of training data is, the more computing time is needed for the training and testing process and the higher accuracy, precision, and recall of the classification are obtained. While the greater the number of k values, the longer computational time needed for the training and testing.

Keywords: *Weighted K-Means Algorithm, Training and Testing data, Fruit Classification, YCbCr, Pattern Recognition.*

*Penulis Korespondensi

1. PENDAHULUAN

Klasifikasi kualitas buah secara terkomputerisasi dengan menggunakan data citra sangat diperlukan sebagai bahan pertimbangan bagi pelaku usaha atau industri buah-buahan pada saat ini. Proses pensortiran yang lebih cepat dan mudah juga dapat meminimalkan kerugian yang ditimbulkan pelaku bisnis dari segi materi. Mengacu pada permasalahan diatas, penelitian ini penting dilakukan, dimana dibutuhkannya suatu sistem yang mudah dan cepat dalam memberikan hasil akurasi yang lebih baik dalam pensortiran buah, agar hasil yang diberikan lebih akurat dari metode pensortiran sebelumnya. Untuk itu pada penelitian ini berusaha mengembangkan metode yang belum pernah digunakan sebelumnya dalam klasifikasi kualitas buah, yaitu metode pengenalan pola yang bernama *weighted k-means*.

Sebelum itu perlu diketahui bahwa *Weighted K-Means* ialah pengembangan dari algoritma *K-Means* yang sudah umum diterapkan dari berbagai pengujian terdahulu. Perbedaan utama *Weighted K-Means* dengan algoritma *k-means* terletak pada penambahan fitur nilai bobot yang digunakan dalam proses

pelatihan dan pengujian[1]. Proses klasifikasi dalam sistem yang dibuat ini dibagi menjadi dua tahapan yaitu tahapan pelatihan dan tahapan pengujian/klasifikasi.

Jenis buah yang digunakan dalam penelitian ini dikhususkan menggunakan buah semangka, agar dapat dilakukan perbandingan dengan penelitian terdahulu yang menggunakan kasus yang sama, sehingga perbandingan akurasi dari metode sebelumnya dengan metode *weighted k-means* ini dapat diketahui mana yang lebih baik untuk digunakan. Pada penelitian ini buah semangka dibagi menjadi tiga kategori kualitas buah yaitu segar, sedang, dan busuk. Dengan terlebih dahulu fitur citra berwarna (RGB) yang ada di citra asli, diekstraksi ke *image* pada ruang warna *YcbCr* sebelum dilakukan proses pengklasifikasian. Untuk hasil akhir dari penelitian ini, parameter tingkat keberhasilan yang digunakan adalah waktu komputasi dan tingkat akurasi. Dari penjabaran diatas dapat ditarik kesimpulan bahwa dalam penelitian tugas akhir kali ini akan membahas pengaruh penerapan algoritma *weighted k-means* dalam klasifikasi tingkat kesegaran buah semangka

2. TINJAUAN PUSTAKA

Klasifikasi kualitas buah menggunakan algoritma SVM[2] telah berhasil dilakukan, namun hanya memiliki tingkat akurasi sebesar 60%. Data citra uji buah semangka yang digunakan menggunakan 3 kategori yaitu, segar, sedang dan busuk. Fitur warna dalam pengujian HSI dan pengimplementasian program menggunakan OpenCV.

Klasifikasi tingkat kematangan buah *strawberry* menggunakan algoritma K-Means[3] telah berhasil dilakukan dengan tingkat akurasi 60%. Jumlah akurasi yang kecil dikarenakan jumlah data *training* yang masih kurang, tapi masih membuktikan bahwa algoritma k-means bisa dilakukan untuk studi kasus pensortiran buah.

Penerapan algoritma *Weighted* K-Means untuk melakukan perbaikan pensortiran dokumen[4] memiliki tingkat akurasi 4-5 % lebih baik dibandingkan dengan algoritma K-Means tradisional. Berdasarkan uji coba yang dilakukan didapatkan hasil yang menunjukkan bahwa penggunaan algoritma *Weighted* K-Means juga dapat dilakukan untuk pensortiran dokumen.

Penelitian klasifikasi dan pengelompokan data berdasarkan tingkat kemiripannya menggunakan algoritma *Weighted* K-Means[5], dari hasil pengujian yang dilakukan didapatkan hasil yang menunjukkan bahwa algoritma *Weighted* K-Means memiliki hasil yang lebih bagus jika dibandingkan dengan algoritma K-Means tradisional.

Klasifikasi 5 jenis penyakit ginjal menggunakan algoritma K-Means[6] didapatkan hasil akhir klasifikasi dengan tingkat akurasi 90%. Parameter uji yang digunakan dalam penelitian ini adalah ureum, kreatinin, dan GFR. Penelitian ini menggunakan 10 data uji yang akan diklasifikasi menggunakan algoritma K-Means.

Prediksi prestasi akademik mahasiswa menggunakan algoritma K-Means[7], dengan hasil pengujian menunjukkan bahwa data mahasiswa berhasil diklasifikasi. Parameter yang digunakan untuk menentukan prestasi akademik mahasiswa adalah nilai IPK.

Jadi dapat ditarik kesimpulan dari berbagai penelitian terdahulu yang menjadi referensi tinjauan pustaka, bahwa tugas akhir ini akan melakukan pengembangan tentang penggunaan algoritma *Weighted* K-Means dalam mengklasifikasi kualitas buah, dalam hal ini dengan studi kasus buah semangka.

3. METODE PENELITIAN

3.1. Alat Bahan

Berikut adalah informasi detail spesifikasi perangkat keras dan perangkat lunak yang digunakan pada penelitian tugas akhir ini antara lain:

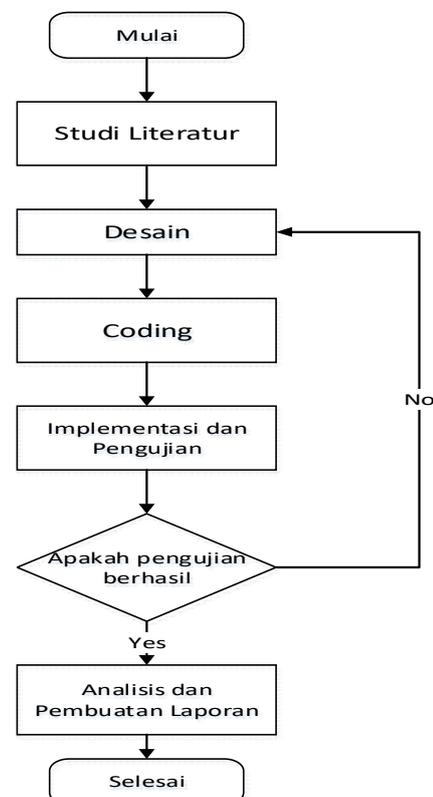
- Laptop Samsung NP300E4X dengan spesifikasi(Intel Inside With NVIDIA Optimus), RAM 2 GB, 300 Gb *Harddisk*, OS Windows 7 (32-bit.).
- Microsoft Office Word, Excel.
- Microsoft Visio
- Adobe Photoshop (sebagai *software* editor foto)
- Matlab 2013r

Bahan yang digunakan yaitu,

- Buah semangka
- Citra buah semangka dengan ukuran 128x128pixel, yang diperoleh dari penelitian metode SVM[1].

3.2. Tahapan Penelitian

Terdapat beberapa tahapan yang akan dilakukan guna mencapai hasil akhir yang mampu menjawab tujuan akhir dari penelitian ini, Gambar 1 yang menampilkan tahap-tahap apa saja yang akan berlangsung dalam tugas akhir ini.



Gambar 1. Diagram alir penelitian

a. Studi Literatur

Studi *literature* merupakan proses pencarian tinjauan pustaka yang berkaitan dengan tugas akhir yang akan dilakukan[8]. Referensi yang menjadi acuan dalam tugas akhir ini berisikan materi tentang ruang warna pada pengolahan citra digital, algoritma K-Means dan Weighted K-Means.

b. Desain

Pada tahap ini dilaksanakan perancangan bagan alir, antar muka dan perancangan fungsi-fungsi dari aplikasi yang akan dibuat untuk melakukan klasifikasi kualitas buah menggunakan algoritma yang telah ditetapkan.

c. Coding

Pada tahap ini akan dilakukan proses coding untuk merealisasikan hasil perancangan menjadi program yang utuh.

d. Implementasi dan Pengujian

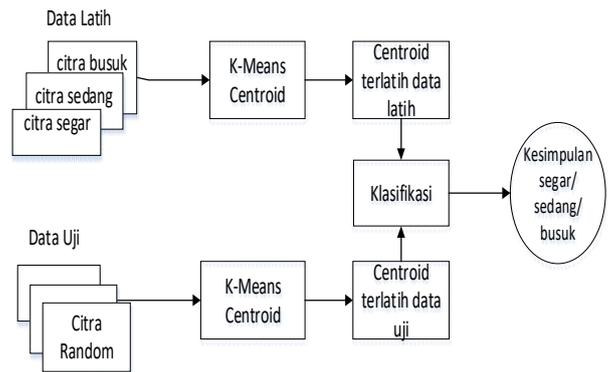
Pada tahap ini akan dilakukan implementasi dan pengujian terhadap program yang telah dibuat untuk mengetahui apakah program atau sistem yang telah dibuat berhasil atau tidak. Apabila program yang dibuat berhasil atau telah sesuai dengan tujuan dan permasalahan yang telah ditetapkan maka akan dilanjutkan ke tahap pembuatan laporan. Dan apabila program yang dibuat gagal atau tidak sesuai dengan tujuan dan rumusan masalah maka proses akan kembali ke tahap desain.

e. Analisa dan Pembuatan Laporan

Tahapan ini akan melakukan analisa dari berbagai uji coba pada program yang telah dilakukan dan melakukan penyusunan laporan dari hasil uji coba dan analisa.

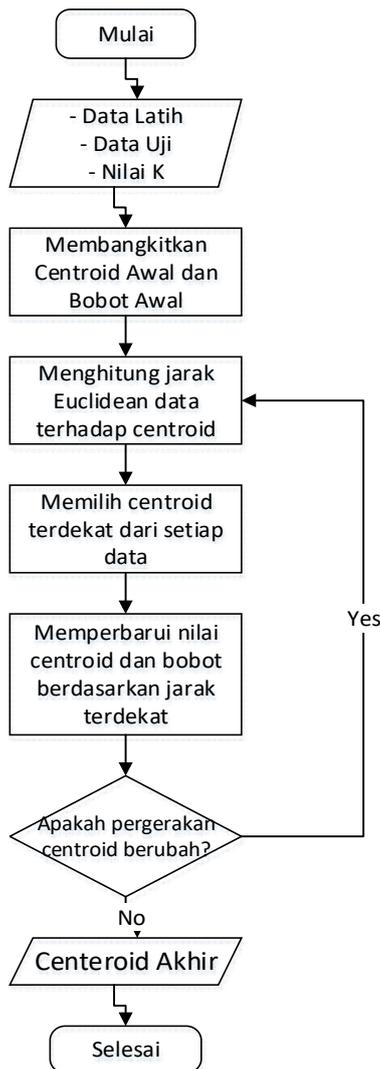
3.3. Metode Klasifikasi dengan *Weighted K-Means*

Metode klasifikasi kualitas buah pada penelitian ini terdiri dari dua tahapan, yaitu pelatihan dan pengujian. Gambar 2 yang merupakan blok diagram pada proses klasifikasi *weighted k-means*.



Gambar 2. Blok diagram klasifikasi *Weighted kmeans*.

Berdasarkan Gambar 2 tahapan-tahapan klasifikasi buah semangka dibagi menjadi 2 yaitu, tahapan pembacaan citra data latih dan dan tahapan pembacaan citra uji. Sehingga akan didapatkan hasil akhir berupa citra dengan kategori segar, sedang maupun busuk. Untuk lebih lengkapnya berikut penjabaran diagram alir yang memperlihatkan langkah-langkah dari algoritma *Wighted K-Means* yang akan digunakan pada penelitian tugas akhir ini.



Gambar 3. Diagram alir Weighted K-Means

Gambar 3 menampilkan tahapan-tahapan penerapan metode *weighted* k-means secara umum. Untuk studi kasus klasifikasi kualitas buah, penjabaran alur sistem ini akan dijelaskan sebagai berikut:

- a. Menyiapkan citra data latih dan uji. Dimana jumlah citra data latih segar, sedang, dan busuk berturut-turut adalah 25, 25, dan 20. Serta jumlah data uji yang tidak diketahui kategorinya sejumlah 70.
- b. Me-*resize* citra semangka yang dipunya ke ukuran 128x128 pixel dengan format (.png) dan (.bmp).
- c. Pembacaan terhadap data latih yang akan digunakan sebagai data acuan dalam melakukan klasifikasi data citra.
- d. Menentukan nilai K, besar kecilnya nilai K akan berpengaruh terhadap hasil klasifikasi. Nilai K merupakan jumlah *cluster*.

- e. Melakukan pembangkitan terhadap centroid awal (titik pusat *cluster*). Ukuran centroid = k x jumlah parameter (R,G,dan B)
- f. Melakukan pembangkitan bobot awal secara acak. Berikut adalah persamaan(1) untuk menghitung bobot dalam *weighted* k-means: Ukuran bobot = 1 x K

$$U_{ij} = W_{ij} \cdot d_{ij} \quad (1)$$

- g. Memilih centroid terdekat dari masing-masing data berdasarkan jarak terbobot.
- h. Meng-*update* posisi centroid berdasarkan jarak terdekat. *Update* dilakukan secara berulang hingga iterasi selesai atau tidak terjadi lagi perubahan nilai centroid. Hal ini berarti centroid terlatih sudah didapatkan.
- i. Tahapan berikutnya adalah melakukan klasifikasi untuk mendapatkan centroid akhir dari data latih. Centroid akhir ini merupakan centroid terlatih yang akan digunakan untuk melakukan klasifikasi data uji.
- j. Melakukan klasifikasi untuk mendapatkan centroid akhir dari data uji, dengan cara yang sama saat tahap pelatihan. Centroid akhir ini merupakan centroid terlatih yang akan digunakan sebagai acuan klasifikasi.
- k. Langkah selanjutnya adalah menghitung jarak *Euclidean* dari centroid data uji terhadap centroid data latih. Berikut ini merupakan persamaan(2) untuk menghitung jarak *euclidean*:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (x_{ir} - x_{jr})^2} \quad (2)$$

dimana :

$d(x_i, x_j)$ = Jarak Euclidiean data uji x_j terhadap data latih x_i

x_{ir} = Data latih x_i ke-r.

x_{jr} = Data uji x_j ke-r.

- l. Langkah terakhir, melakukan pelabelan data uji berdasarkan jarak centroid terdekat, agar dapat ditarik kesimpulan citra uji tersebut masuk ke kategori buah semangka segar, sedang ataupun busuk.

3.4. Metode Pengujian

Pengujian ini dilakukan dengan menguji citra semangka menggunakan beberapa resolusi citra diresize hingga ukuran 128x128 piksel. Pengujian untuk masing-masing data latih dan data uji menggunakan fitur warna YCbCr, dimana sebelumnya fitur warna RGB dikonversi kedalam YcbCr.

Berikut ini adalah beberapa parameter dalam proses pengujian yang akan berlangsung di tugas akhir ini:

- a. Pengaruh jumlah data citra latih (*sample*) buah semangka. Pengujian terhadap seberapa baik akurasi data uji dalam proses klasifikasi berdasarkan banyaknya ketersediaan data latih (*sample*) yang disiapkan dalam proses *clustering*. Pada penelitian ini terdapat dua kumpulan *dataset*, yaitu data latih dan data uji, dengan perincian seperti dijelaskan di Tabel 1, dibawah ini:

TABEL 1. PEMBAGIAN DATASET

Data	Citra Semangka			Total
	Segar	Sedang	Busuk	
Latih	25	25	20	70
Uji	25	25	20	70

- b. Pengaruh nilai k terhadap akurasi. Pengujian ini dapat dilakukan dengan parameter jumlah nilai k berbeda yaitu 2,3,4 atau lebih, guna menentukan persentasi tingkat akurasi, *precision* maupun *recall*. K pada metode weighted k-means merupakan pusat *cluster*.

Evaluasi akhir hasil klasifikasi dapat dilakukan dengan menghitung nilai *Precision*, *Recall* dan *Accuracy*. Untuk lebih jelasnya maka persamaan untuk menghitung masing-masing parameter pengujian dikaji pada Persamaan (3), (4) dan (5).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{5}$$

Dimana :

- TP (True Positive) adalah banyaknya hasil klasifikasi benar untuk suatu kelas yang bernilai positive.
- TN (True Negative) adalah banyaknya hasil klasifikasi benar untuk suatu kelas yang bernilai negative.
- FP (False Positive) adalah banyaknya hasil klasifikasi salah untuk suatu kelas dan bernilai negative.
- FN (False Negative) adalah banyaknya hasil klasifikasi salah untuk suatu kelas dan bernilai positive.

4. HASIL DAN PEMBAHASAN

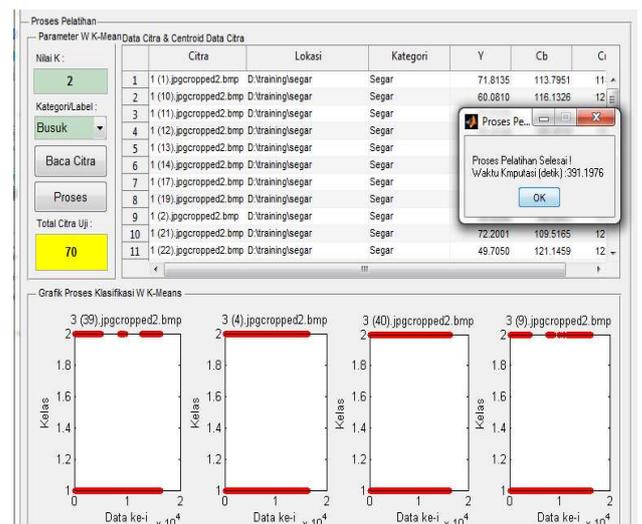
4.1. Pengujian Program

Hasil pengujian proses klasifikasi kualitas buah semangka akan dibahas dalam dua poin utama. Pertama berdasarkan pengaruh data citra uji terhadap jumlah data latih berbeda-beda. Selain itu parameter pengujian juga dilakukan terhadap nilai K yang bervariasi.

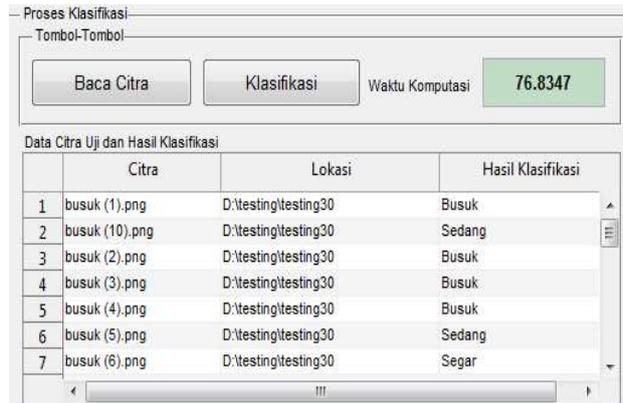
4.1.1. Pengujian berdasarkan parameter jumlah data latih yang bervariasi

- a. Pengujian 30 citra data uji dengan jumlah data citra latih = 30. Pengujian ini dilakukan terhadap data latih dengan jumlah 30 yang terdiri atas 10 data citra berkualitas segar, 10 data citra berkualitas sedang, dan 10 data citra berkualitas buruk. Gambar 4 yang merupakan tampilan untuk proses pelatihan pada sistem. Adapun jumlah data uji yang digunakan adalah 30 data uji, Gambar 5 menampilkan proses pengujian/ klasifikasi dari sistem. Didapatkan hasil klasifikasi benar sebanyak 5 citra uji. Adapun waktu komputasi untuk proses pelatihan yaitu 49.9727 detik, sedangkan waktu komputasi untuk proses klasifikasi yaitu 8.5467 detik.

$$Akurasi = \frac{Juml\ tebak\an\ benar \times 100}{juml\ data\ uji} = \frac{5 \times 100}{30} = 50\%$$



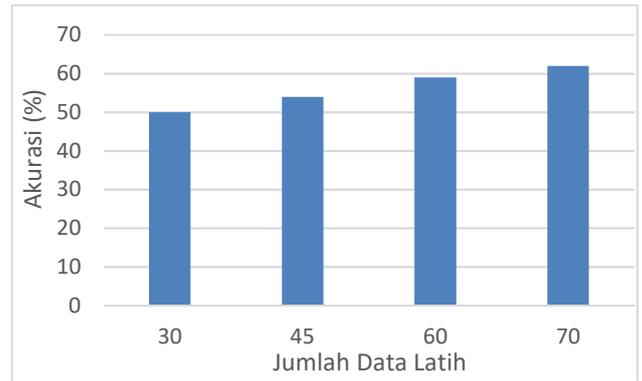
Gambar 4. Proses pelatihan untuk mendapat centroid data latih



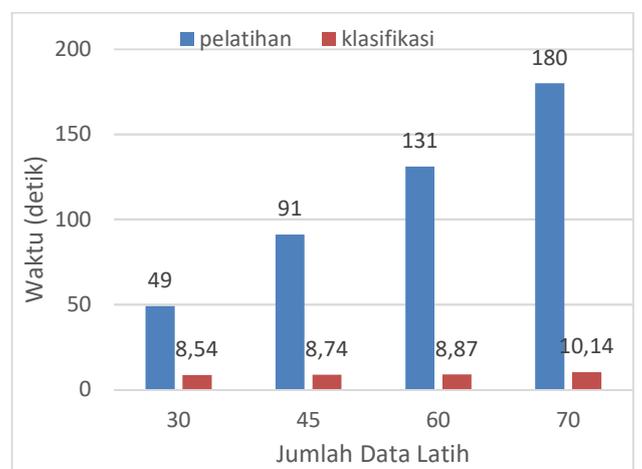
Gambar 5. klasifikasi semangka *weighted kmeans*

- b. Pengujian 30 citra data uji dengan data Citra Latih=45. Pengujian ini dilakukan terhadap data latih dengan jumlah 45 yang terdiri atas 15 data citra berkualitas segar, 15 data citra berkualitas sedang, dan 15 data citra berkualitas buruk. Adapun jumlah data uji yang digunakan adalah 30 data uji. Didapatkan hasil klasifikasi yang benar 5, dengan tingkat akurasi sebesar 50%. Adapun waktu komputasi untuk proses pelatihan yaitu 91.4886 detik, sedangkan waktu komputasi untuk proses klasifikasi yaitu 8.7491 detik.
- c. Pengujian citra data uji dengan data citra latih=60. Pengujian ini dilakukan terhadap data latih dengan jumlah 60 yang terdiri atas 20 data citra berkualitas segar, 20 data citra berkualitas sedang, dan 20 data citra berkualitas buruk. Adapun jumlah data uji yang digunakan adalah 30 data uji. Didapatkan hasil klasifikasi dengan tingkat akurasi sebesar 30%. Adapun waktu komputasi untuk proses pelatihan yaitu 131.9597 detik, sedangkan waktu komputasi untuk proses klasifikasi yaitu 8.879 detik.

Hasil akhir keseluruhan yang memperlihatkan hubungan atau pengaruh jumlah data latih terhadap tingkat akurasi, serta waktu komputasi proses pelatihan dan proses klasifikasi, dapat dilihat dalam grafik.



Gambar 6. Grafik pengaruh jumlah data latih terhadap tingkat akurasi.



Gambar 7. Grafik pengaruh jumlah data latih terhadap waktu komputasi

Berdasarkan empat uji coba yang dilakukan terhadap jumlah data latih yang bervariasi, dimana jumlah data uji dan parameter *k* bernilai tetap, didapatkan hasil yang menunjukkan bahwa semakin besar jumlah data latih, maka waktu komputasi yang dibutuhkan untuk proses pelatihan dan pengujian akan semakin meningkat. Sedangkan tingkat akurasi hasil klasifikasi yang di dapatkan juga akan semakin bagus.

4.1.2 Pengujian citra data uji berdasarkan parameter nilai *K* yang bervariasi

- a. Pengujian citra uji Dengan Nilai *K* = 2
 Pengujian ini dilakukan dengan parameter nilai *k* = 2, terhadap data latih yang terdiri atas 10 data citra berkualitas segar, 10 data citra berkualitas sedang, dan 10 data citra berkualitas buruk. Adapun jumlah data uji yang digunakan adalah 10 data uji. didapatkan hasil klasifikasi dengan tingkat akurasi sebesar 60%. Adapun waktu komputasi untuk proses pelatihan yaitu

74.32 detik, sedangkan waktu komputasi untuk proses klasifikasi yaitu 12.17 detik.

$$\text{Akurasi} = \frac{\text{Jumlah klasifikasi benar} \times 100}{\text{jumlah data uji}} = \frac{6 \times 100}{10} = 60\%$$

b. Pengujian citra uji Dengan Nilai K = 3

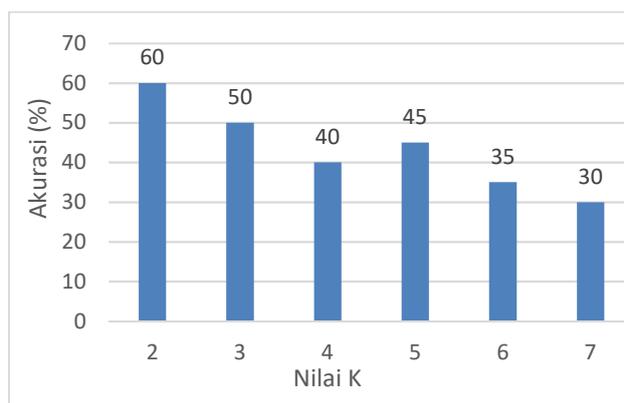
Pengujian ini dilakukan dengan parameter nilai k = 3, terhadap data latih yang terdiri atas 10 data citra berkualitas segar, 10 data citra berkualitas sedang, dan 10 data citra berkualitas buruk. Adapun jumlah data uji yang digunakan adalah 10 data uji. Adapun didapatkan hasil klasifikasi dengan tingkat akurasi sebesar 50%. Waktu komputasi untuk proses pelatihan yaitu 142,32 detik, sedangkan waktu komputasi untuk proses klasifikasi yaitu 21,78 detik.

$$\text{Akurasi} = \frac{\text{Jumlah klasifikasi benar} \times 100}{\text{jumlah data uji}} = \frac{5 \times 100}{10} = 50\%$$

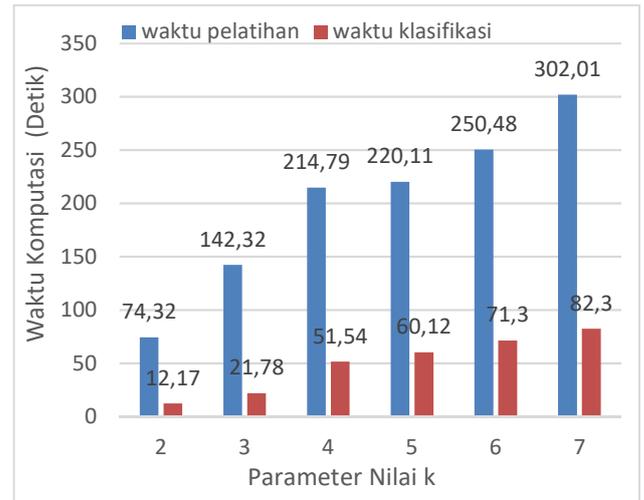
TABEL II. CONFUSION MATRIX DENGAN 30 DATA UJI RANDOM

No	Kategori	Hasil Klasifikasi Benar		
		Segar	Sedang	Busuk
1	Segar	4	2	4
2	Sedang	1	4	5
3	Busuk	2	3	5

Adapun grafik yang memperlihatkan hubungan atau pengaruh jumlah nilai k terhadap tingkat akurasi dan waktu komputasi proses pelatihan dan proses klasifikasi, dapat dilihat dalam grafik.



Gambar 8. Grafik perubahan nilai k terhadap akurasi.



Gambar 9. Grafik perubahan waktu komputasi terhadap nilai K.

Berdasarkan uji coba yang dilakukan terhadap pengaruh parameter nilai k yang bervariasi, dengan jumlah data uji tetap, didapatkan hasil yang menunjukkan bahwa semakin besar jumlah nilai k maka waktu komputasi yang dibutuhkan untuk proses pelatihan dan pengujian akan semakin meningkat. Tetapi tingkat akurasi hasil klasifikasi yang didapatkan semakin kecil. Semakin kecilnya tingkat akurasi yang diperoleh dimungkinkan karena karakteristik data citra latih yang mirip satu dengan yang lainnya pada masing-masing kategori.

4.1.3 Pengujian nilai *precision* dan *recall*

Pada studi kasus pengujian buah semangka menggunakan algoritma *weighted k-means*, *Recall* disini mengacu banyaknya citra semangka yang terdeteksi benar untuk salah satu kategori namun seharusnya tidak. Sebaliknya *Precision* disini merupakan banyaknya citra semangka yang terdeteksi tidak benar untuk salah satu kategori yang seharusnya benar. Hasil implementasi:

Pengujian dilakukan terhadap data latih berjumlah 70 dan data uji 30, serta nilai k=3. Didapatkan hasil sebagai berikut.

Waktu komputasi pelatihan = 391,1976 detik
 Waktu komputasi klasifikasi = 76,8347 detik.

$$\text{Precision data segar} = \frac{4}{4+1+2} \times 100\% = 57\%$$

$$\text{Precision data sedang} = \frac{4}{2+4+3} \times 100\% = 44\%$$

$$\text{Precision data busuk} = \frac{5}{4+5+5} \times 100\% = 35\%$$

Selanjutnya mengambil nilai rata-rata presicion dari 3 kategori kematangan buah semangka.

$$\text{Precision}_{70} = \frac{57+44+35}{3} = 45,3\%$$

$$\text{Recall data segar} = \frac{4}{4+2+4} \times 100\% = 40\%$$

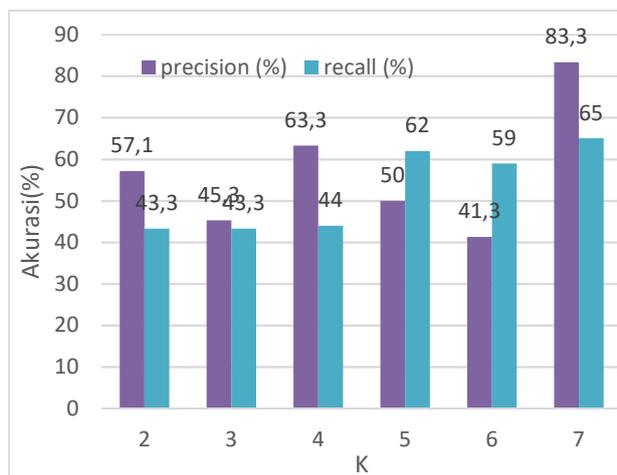
$$\text{Recall data sedang} = \frac{4}{1+4+5} \times 100\% = 40\%$$

$$\text{Recall data busuk} = \frac{5}{2+3+5} \times 100\% = 50\%$$

Selanjutnya mengambil nilai rata-rata recall dari 3 kategori kematangan buah semangka.

$$\text{Recall}_{70} = \frac{40+40+50}{3} = 43,3\%$$

Setelah didapatkan hasil akhir, pengujian kemudian dilanjutkan menggunakan cara yang sama untuk menemukan nilai *precision* dan *recall* dengan pengaruh jumlah data latih bervariasi 30, 45, 60, 70. Dan dengan nilai *k* berjumlah 2 dan 3. Untuk hasil hasil pengujian keseluruhan, terdapat dalam Gambar 10 yang memperlihatkan grafik presentasi perubahan dari *precision* dan *recall*.



Gambar 10. Grafik hasil precision dan recall

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil pengujian klasifikasi kualitas buah semangka dengan algoritma *weighted kmeans* yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut:

1. Berdasarkan pengaruh jumlah data latih, nilai presentasi tingkat akurasi yang tertinggi adalah 62%, dengan jumlah data latih sebanyak 70. Sedangkan presentasi terendah sebesar 50% dengan jumlah data latih 30.
2. Berdasarkan pengaruh nilai *k*, nilai presentasi tingkat akurasi yang tertinggi adalah 60%, dengan nilai *k*=2. Sedangkan terendah sebesar 30% dengan *k*=7.

3. Nilai *precision* dan *Recall* tertinggi yang didapat adalah 83,3% dan 65% dengan jumlah data latih 70 dan *k*=2. Sedangkan yang terendah didapat adalah 35% dan 43,3%.
4. Banyaknya jumlah citra latih/uji dan nilai *k* mempengaruhi waktu komputasi yang dibutuhkan untuk proses pelatihan dan pengujian, dimana durasi waktu komputasi proses pelatihan dan pengujian akan semakin meningkat.

5.2. Saran

Untuk mendapatkan hasil akurasi yang lebih baik dan optimal dalam penerapan algoritma *weighted k-means*, di penelitian selanjutnya diharapkan menggunakan studi kasus dengan jenis buah yang berbeda, dikarenakan citra semangka memiliki karakteristik warna yang hampir mirip di kulitnya dari masing-masing kategori segar, sedang maupun busuk.

DAFTAR PUSTAKA

- [1] A. M. Baswade, K. D. Joshi, and P. S. Nalwade, "A Comparative Study Of K-Means And Weighted K-Means For Clustering," *Int. J. Eng. Res. Technol.*, vol. 1, no. 10, pp. 1–4, 2012.
- [2] F. Fadilah, I. G. P. S. Wijaya, and F. Bimantoro, "Sistem Deteksi Kualitas Semangka Berdasarkan Pendaran Warna dan Menggunakan Support Vector Machine (SVM)," *Dielektrika*, vol. 5, no. 2, pp. 142–148, 2018.
- [3] O. H. Indrianto, "Klasifikasi Kematangan Buah Strawberry Menggunakan Algoritma K-Means," Universitas Mataram, 2016.
- [4] M. Mandloi and A. Kothari, "An Improved Document Clustering Approach Using Weighted K-Means Algorithm," *IJCSN Int. J. Comput. Sci. Netw.*, vol. 6, no. 2, pp. 2277–5420, 2017.
- [5] K. Kerdprasop, N. Kerdprasop, and P. Sattayatham, "Weighted K-Means for Density-Biased Clustering," *Int. Conf. Data Warehous. Knowl. Discov.*, pp. 488–489, 2005.
- [6] A. S. Irtawaty, "Klasifikasi Penyakit Ginjal dengan Metode K-Means," *J. Teknol. Terpadu*, vol. 5, no. 1, pp. 3–7, 2017.
- [7] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, "Application of k-Means Clustering algorithm for prediction of Students' Academic Performance," *Int. J. Comput. Sci. Inf. Secur.*, vol. 7, pp. 292–295, 2010.
- [8] T. Sutanto, "Kombinasi Penyaring Warna Kulit Kplisit pada Bidang Warna RGB Dan YCbCr Untuk Meningkatkan Akurasi Sistem Pendeteksi Warna Kulit," STIKOM Surabaya, 2010