

# Penentuan Mutu Kelapa Sawit Menggunakan Metode K-Means Clustering

Andri Nofiar Am<sup>1</sup>, Sarjon Defit<sup>2</sup>, Sumijan<sup>3</sup>

<sup>1</sup>Politeknik Kampar

<sup>2,3</sup> Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK

E-mail: Andrinofiar90@gmail.com

## Abstract

The classification of the quality of palm oil in PT Tasma Puja is still done by laboratory testing and then the data is saved manually in Excel. The method of grouping takes time and allows data to be lost. With the development of knowledge, it can be replaced by a data mining approach that can be used to classify the quality of palm oil based on its standards. The k-Means clustering method can be applied to classify the quality of palm oil based on water, dirt and free fatty acids. The data used is the quality data of palm oil in December 2017 as many as 31 data with criteria of good, very good and not good. The test results contained 3 clusters, namely cluster 0 for good categories amounted to 12 data, cluster 1 for very good category amounted to 13 data and cluster 2 for less good categories amounted to 6 data. The k-Means clustering method can be used for data processing using the concept of data mining in grouping data according to criteria.

**Kata kunci:** *Data Mining, K-Means Clustering, Cluster*

## Abstrak

Pengelompokan mutu minyak kelapa sawit pada PT Tasma Puja masih di lakukan dengan cara uji laboratorium dan selanjutnya data disimpan secara manual di *excel*. Cara pengelompokan tersebut membutuhkan waktu dan memungkinkan data hilang. Dengan perkembangan ilmu pengetahuan nantinya bisa di gantikan dengan pendekatan *data mining* yang dapat digunakan dalam mengelompokan mutu minyak kelapa sawit berdasarkan standarnya. Metode *k-Means clustering* dapat diterapkan untuk mengelompokan suatu kualitas minyak kelapa sawit berdasarkan air, kotoran dan asam lemak bebas. Data yang digunakan adalah data kualitas minyak kelapa sawit pada bulan desember 2017 sebanyak 31 data dengan kriteria baik, sangat baik dan kurang baik. Hasil pengujian terdapat 3 *cluster* yaitu *cluster* 0 kategori baik berjumlah 12 data, *cluster* 1 kategori sangat baik berjumlah 13 data dan *cluster* 2 kategori kurang baik berjumlah 6 data. Metode *k-Means clustering* dapat digunakan untuk proses pengolahan data menggunakan konsep *data mining* dalam mengelompokkan data sesuai kriteria.

**Kata kunci:** *Data Mining, K-Means Clustering, Cluster*

## 1. Pendahuluan

Indonesia merupakan negara yang memiliki banyak lahan pertanian dan perkebunan, salah satunya yaitu kelapa sawit. Salah satu tumbuhan industri penghasil minyak adalah kelapa sawit, baik itu minyak masak, minyak industri, minyak bahan bakar (biodiesel). Jumlah total luas area perkebunan kelapa sawit di Indonesia pada saat ini mencapai 11,9 juta hektar berdasarkan badan pusat statistik.

Untuk menghitung kualitas minyak kelapa sawit dilakukan uji laboratorium berdasarkan kandungan air, kotoran dan k asam lemak bebas. Salah satu faktor yang mempengaruhi dalam

Dikirim : 2019-03-14

Diterima : 2019-03-20

Diterbitkan : 2019-04-01

DOI : <https://doi.org/10.29165/komtekinf.v5i2>

kualitas minyak kelapa sawit yaitu kadar Asam Lemak Bebas (ALB). Dimana tandan buah kelapa sawit ini ketika panen kadar ALB nya ini akan meningkat 0.1% setiap hari, akan tetapi kadar ini tidak boleh lebih dari 2-3% ketika masuk di pabrik, karena bobot ALB yang baik hanya 3% [1].

Proses penyimpanan data air, kotoran dan asam lemak bebas yang dihasilkan setiap kali produksi dari hasil uji laboratorium masih disimpan secara manual menggunakan aplikasi *microsoft excel*. Dengan penyimpanan data tersebut terdapat kelemahan yaitu susahnya dalam melakukan penilaian kualitas produksi minyak kelapa sawit yang dihasilkan. Produksi minyak kelapa sawit yang dilakukan setiap hari pada PT. Tasma Puja menyebabkan banyaknya penyimpanan data yang berdampak pada lamanya waktu yang dibutuhkan untuk mencari hasil terbaik dari setiap kali produksi minyak kelapa sawit.

Dengan perkembangan ilmu pengetahuan yang semakin pesat nantinya, proses penyimpanan data hasil produksi pada PT. Tasma Puja dapat digantikan dengan pendekatan *data mining* yang digunakan dalam mengelompokkan kualitas minyak kelapa sawit berdasarkan bobot yang telah ditetapkan, dengan melakukan pembagian data kedalam kelas atau *cluster* berdasarkan tingkat kesamaannya. Sehingga Data yang diolah dengan menggunakan *data mining* menghasilkan pengetahuan atau *Knowledge* yang baru, dengan bersumber dari data lama hasil produksi minyak kelapa sawit.

## 2. Tinjauan Literatur

*Knowledge Discovery in Database* (KDD) merupakan suatu metode yang digunakan untuk menganalisis data pada penerapan *data mining* untuk mencari pengetahuan, dan informasi yang lebih dari sebuah kumpulan data. *Knowledge Discovery in Database* (KDD) yaitu suatu kegiatan yang terdiri dari beberapa tahapan, seperti *Cleaning and Integration*, *Selection and Transformation*, *Data mining*, dan *Evaluation and Interpretation*. Pada proses *Evaluation and Interpretation* ini merupakan cakupan dari proses pemeriksaan baik pola itu atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa sebelumnya. Dalam memilih suatu metode atau *algoritma* yang digunakan sangat mempengaruhi dari proses dan tujuan dari *Knowledge Discovery in Database* (KDD) [2].

*Data mining* merupakan suatu proses penggalian pola dari sekumpulan data yang besar dengan menggunakan teknik tertentu yang menghasilkan nilai lebih, baik itu pengetahuan dan informasi dari data yang belum diketahui. *Data mining* dapat melakukan satu atau lebih dari teknik pembelajaran komputer, dimana proses tersebut dapat dilakukan secara otomatis ataupun semi otomatis. Pola yang dicari menghasilkan nilai tambah ekonomi dari hasil yang ditemukan [3].

Salah satu teknik pengelompokkan dari data mining adalah *clustering*. *Clustering* adalah pengelompokkan sejumlah data atau objek ke dalam *cluster* sehingga setiap dalam *cluster* tersebut berisi data yang semirip mungkin dan berbeda dari objek dalam *cluster* lainnya. *Clustering* merupakan salah satu teknik pengelompokkan dalam data mining yaitu pengelompokkan sejumlah data ataupun objek ke dalam suatu kelompok atau *cluster* atau *group* sehingga setiap data dalam *cluster* tersebut berisi suatu data yang memiliki kemiripan ataupun berbeda dengan objek dalam *cluster* yang lain [4].

Algoritma *K-Means* adalah pengelompokkan data yang berdasarkan titik pusat dari klaster (*centroid*) yang terdekat dengan data yang merupakan algoritma klasterisasi. Adapun tujuan dari algoritma ini yaitu mengelompokkan data dengan memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antar klaster. Fungsi jarak merupakan ukuran kemiripan yang digunakan dalam klaster, sehingga hasil yang maksimal dari kemiripan data berdasarkan jarak terpendek antara data terhadap titik *centroid* [5].

Algoritma *K-Means* merupakan suatu metode yang digunakan untuk menemukan kelompok dari dokumen atau objek yang *non-overlapping*, dan juga dapat disebut sebagai suatu *algoritma* yang paling efektif untuk mengelompokkan kumpulan data. Algoritma ini dapat diimplementasikan, dijalankan dan diproses dengan cepat, mudah beradaptasi [6].

Adapun langkah-langkah perhitungan menggunakan algoritma *K-Means* ini yaitu :

1. Masukkan jumlah  $K$  cluster,  $K$  disini berarti banyaknya cluster yang ingin dibentuk.
2. Tentukan nilai awal centroid ( $K$  pusat cluster) untuk  $k$  cluster dari sejumlah dataset yang bisa dilakukan dengan berbagai cara yang salah satunya dengan secara acak atau *random*.
3. Semua data/objek dialokasikan ke cluster yang terdekat antara dua objek yang sesuai dengan jarak kedua objek tersebut. Sama halnya dengan kedekatan satu data ke cluster tertentu. Langkah ini dapat dihitung dari jarak tiap data ke tiap pusat cluster (*centroid cluster*). Untuk menghitung setiap titik pada cluster terdekat bisa menggunakan rumus jarak *Euclidean*, sebagai berikut :

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Dimana:

$D_{(i,j)}$  = jarak data ke  $i$  ke pusat cluster  $j$

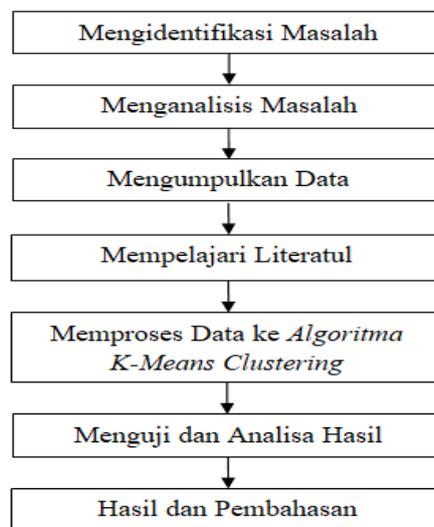
$X_{ki}$  = Data ke  $i$  pada atribut data ke  $k$

$X_{kj}$  = Titik pusat ke  $j$  pada atribut ke  $k$

4. Hitung ulang untuk mencari nilai centroid dari setiap  $k$  cluster dengan jumlah data  $m$  untuk menemukan nilai centroid  $K$  cluster yang baru. Adapun titik pusat cluster (*centroid cluster*) yaitu rata-rata dari semua titik dalam cluster. Adapun cara mencarinya bisa menggunakan *median* dari data kelompok tersebut [7].

### 3. Metodologi Penelitian

Metodeologi penelitian disini dijabarkan dengan kerangka kerja penelitian. Kerangka kerja merupakan urutan atau uraian alur kerja sistem yang harus diikuti. Penelitian ini akan dilakukan secara berurutan dari awal penelitian sampai hasil yang akan didapatkan dengan menggunakan metode *K-Means clustering*. Adapun kerangka kerja yang harus diikuti untuk penelitian ini dapat dilihat seperti gambar 2 sebagai berikut.



Gambar 1 : Kerangka Kerja Penelitian

Keterangan :

1. Mengidentifikasi Masalah  
Masalah yang diidentifikasi dalam penelitian ini adalah mengelompokkan asam lemak bebas untuk menentukan kualitas minyak kelapa sawit yang dihasilkan pada PT. Tasma Puja.
2. Menganalisis Masalah  
Analisis masalah adalah untuk dapat memahami masalah yang telah ditentukan ruang lingkup dan batasannya. Dengan menganalisa masalah yang telah ditentukan tersebut, maka diharapkan

masalah dapat dipahami dengan baik. Proses pengambilan keputusan dalam pengelompokan kualitas minyak kelapa sawit pada PT. Tasma Puja.

### 3. Mengumpulkan data

Untuk memperoleh data-data yang dibutuhkan dalam proses penulisan tesis ini, maka penulis menggunakan metode-metode sebagai berikut :

#### a. Studi Kepustakaan (*Library Research*)

Studi kepustakaan adalah suatu metode pengumpulan data yang dilakukan pada perpustakaan yang harus diperhatikan sumber yang diteliti yaitu buku bacaan yang menjadi pembahasan sumber masalah yang dibahas, dan bahan-bahan yang berhubungan dengan masalah yang diteliti.

#### b. Pengamatan (*Observation*)

Untuk mengumpulkan informasi dilakukan pengamatan langsung pada laboratorium yang bertujuan untuk mendapatkan keterangan yang berhubungan dengan objek yang akan diteliti.

#### c. Dokumentasi

Dokumentasi merupakan cara memperoleh data dan informasi yang dibutuhkan dengan mengumpulkan data dari dokumen perusahaan yang berkaitan dengan proses pengumpulan data.

### 4. Mempelajari Literatur

Setelah mengumpulkan data yang akan di proses dalam studi literatur ini, terlebih dahulu mengetahui pengetahuan atau *knowledge* yang akan diterapkan dalam metode ini, literatur yang akan dipelajari ini bersumber dari jurnal-jurnal ilmiah yang di publikasikan di internet.

### 5. Memproses Data Ke Algoritma *K-Means Clustering*

Setelah melakukan pengumpulan data dan mempelajari literatur, selanjutnya dilakukan proses ke algoritma *K-Means Clustering* yaitu :

#### a. Menentukan jumlah *cluster*

#### b. Menentukan titik pusat *centroid* secara *random*

#### c. Menghitung jarak terdekat *centroid*

#### d. Mengelompokkan data ke masing-masing *cluster*

### 6. Menguji Dan Analisa Hasil

Penulis melakukan pelatihan dan pengujian setelah melakukan pengumpulan data, proses ke algoritma *K-Means clustering*, dan penginstalan *software* yang akan digunakan yaitu *RipedReader*, untuk selanjutnya melakukan pelatihan dan pengujian terhadap data. Pengujian adalah data yang diperoleh dari pemrosesan data awal dengan menggunakan metode *K-Means clustering*. Hasil yang diperoleh dari pengujian kemudian akan dianalisa agar kondisi lebih akurat.

### 7. Hasil Dan Pembahasan

Kesimpulan hasil dari pengelompokan kualitas minyak kelapa sawit berdasarkan air, kotoran dan asam lemak bebas menjadi baik, sangat baik dan kurang baik, dengan menggunakan metode *K-Means clustering* dari tahapan proses menguji dan analisa hasil, sehingga menghasilkan kesimpulan yang baik. Untuk mengetahui apakah model yang dirancang tersebut sesuai dengan yang diharapkan.

## 4. Hasil dan Diskusi

Dari *Knowledge Discovery in Database (KDD)*, algoritma *K-Means Clustering* dapat kita pahami bahwa langkah-langkah penyelesaiannya dapat dijelaskan sebagai berikut :

### 1. Proses Seleksi Data

Dalam proses *Knowledge Discovery in Database (KDD)* seleksi data merupakan proses himpunan data dan menciptakan data target pada sampel data, dimana penemuan akan dilakukan selanjutnya hasil seleksi disimpan dalam suatu berkas yang terpisah dari data basis data operasional. Setelah data didapatkan dari pihak PT. Tasma Puja seperti data awal diatas maka dilakukan proses seleksi data, proses ini dilakukan agar memudahkan proses perhitungan

algoritma *K-Means Clustering*. Adapun data yang digunakan adalah data pada bulan Desember 2017 yang dapat dilihat pada tabel sebagai berikut :

**Tabel 1** : Daftar Seleksi Data Bulan Desember

NO	Air	Kotoran	ALB
1	0.280	0.019	3.610
2	0.000	0.000	0.000
3	0.000	0.000	0.000
4	0.320	0.019	3.520
5	0.310	0.019	3.750
6	0.260	0.019	3.200
7	0.340	0.019	3.370
8	0.320	0.019	3.510
9	0.290	0.010	3.840
10	0.000	0.000	0.000
11	0.230	0.019	4.340
12	0.350	0.019	4.040
13	0.350	0.018	3.810
14	0.310	0.018	3.600
15	0.450	0.019	3.990
16	0.340	0.019	3.710
17	0.000	0.000	0.000
18	0.280	0.019	3.560
19	0.290	0.020	4.160
20	0.240	0.019	3.460
21	0.300	0.019	3.940
22	0.310	0.019	3.960
23	0.270	0.019	4.120
24	0.270	0.019	4.310
25	0.000	0.000	0.000
26	0.240	0.019	4.790
27	0.300	0.019	4.190
28	0.270	0.019	4.080
29	0.300	0.018	4.540
30	0.320	0.018	4.310
31	0.000	0.000	0.000

## 2. Input Data

Data yang diinputkan adalah data yang telah didapat dari PT. Tasma Puja bulan desember tahun 2017 yang dapat di lihat pada tabel 1 diatas. Sebelum dilakukan input data kesistem harus dianalisa dan melakukan proses perhitungan secara benar, adapun proses dari algoritma *K-Means Clustering* adalah sebagai berikut :

### a. Menentukan jumlah *cluster* (*K*)

K disini adalah sebagai jumlah *cluster* yang akan dibentuk, pada penelitian ini dibentuk 3 jenis *cluster* yakni air, kotoran dan asam lemak bebas yang didapat dari rekapitulasi data bulanan mutu CPO PT. Tasma Puja yang diambil pada bulan Desember 2017, *cluster* tersebut terbagi menjadi baik, sangat baik dan kurang baik.

### b. Menentukan Titik Pusat Centroid Secara Random

Setelah menentukan jumlah *cluster* sebanyak 3 *cluster*, maka selanjutnya ditentukanlah titik pusat secara *random* disini diambil untuk Pusat *cluster* 0 ( $C_0$ ) : di ambil dari nomor 20 untuk Pusat *cluster* 1 ( $C_1$ ) : di ambil dari nomor 26 dan Pusat *cluster* 2 ( $C_2$ ) : di ambil dari nomor 6 dapat dilihat dari tabel nilai *random* sebagai berikut :

**Tabel 2** : Nilai Centroid Random

Centroid	Air	Kotoran	ALB
$C_0$	0,240	0,019	3,460

C <sub>1</sub>	0,240	0,019	4,790
C <sub>2</sub>	0,260	0,019	3,200

c. Menghitung Jarak Terdekat *Centroid*

Rumus jarak *euclidean distance* digunakan untuk menghitung jarak-jarak objek kepusat terhadap masing-masing *centroid* sehingga ditemukan jarak yang paling terdekat dari setiap data dengan *centroid*. Berikut rumus persamaan Mengelompokkan data masing-masing *cluster* Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* jarak terdekat. Adapun proses hitungnyanya sebagai berikut :

1) Proses Iterasi I (Pertama)

Untuk menghitung pusat *cluster* 0, *cluster* 1 dan *cluster* 2 pada tahapan ini akan dihitung jarak setiap data kemasing-masing *centroid* menggunakan rumus *Euclidean Distance* (D) iterasi pertama. Menghitung jarak masing-masing data ke titik pusat sebagai berikut :

1. Proses Hitung *cluster* (C<sub>0</sub>)

a. Pada *stage* (D1), jarak dengan titik pusat *cluster* (C<sub>0</sub>) yaitu :

$$D_1 = \sqrt{(1_x - C_{1x})^2 + (1_y - C_{1y})^2 + (1_z - C_{1z})^2}$$

$$= \sqrt{(0,280 - 0,240)^2 + (0,019 - 0,019)^2 + (3,610 - 3,460)^2}$$

$$= \mathbf{0,155}$$

b. Pada *stage* (D2), jarak dengan titik pusat *cluster* (C<sub>0</sub>) yaitu :

$$D_1 = \sqrt{(1_x - C_{1x})^2 + (1_y - C_{1y})^2 + (1_z - C_{1z})^2}$$

$$= \sqrt{(0,0 - 0,240)^2 + (0,0 - 0,019)^2 + (0,0 - 3,460)^2}$$

$$= \mathbf{3,468}$$

2. Proses Hitung *cluster* (C<sub>1</sub>)

a. Pada *stage* (D1), jarak dengan titik pusat *cluster* (C<sub>1</sub>) yaitu :

$$D_1 = \sqrt{(1_x - C_{2x})^2 + (1_y - C_{2y})^2 + (1_z - C_{2z})^2}$$

$$= \sqrt{(0,280 - 0,240)^2 + (0,019 - 0,019)^2 + (3,610 - 4,790)^2}$$

$$= \mathbf{1,181}$$

b. Pada *stage* (D2), jarak dengan titik pusat *cluster* (C<sub>1</sub>) yaitu :

$$D_2 = \sqrt{(2_x - C_{2x})^2 + (2_y - C_{2y})^2 + (2_z - C_{2z})^2}$$

$$= \sqrt{(0,0 - 0,240)^2 + (0,0 - 0,019)^2 + (0,0 - 4,790)^2}$$

$$= \mathbf{4,796}$$

3. Proses Hitung *cluster* (C<sub>2</sub>)

a. Pada *stage* (D1), jarak dengan titik pusat *cluster* (C<sub>2</sub>) yaitu :

$$D_1 = \sqrt{(1_x - C_{0x})^2 + (1_y - C_{0y})^2 + (1_z - C_{0z})^2}$$

$$= \sqrt{(0,280 - 0,260)^2 + (0,019 - 0,019)^2 + (3,610 - 3,200)^2}$$

$$= \mathbf{0,410}$$

b. Pada *stage* (D2), jarak dengan titik pusat *cluster* (C<sub>2</sub>) yaitu :

$$D_1 = \sqrt{(1_x - C_{0x})^2 + (1_y - C_{0y})^2 + (1_z - C_{0z})^2}$$

$$= \sqrt{(0,0 - 0,260)^2 + (0,0 - 0,019)^2 + (0,0 - 3,200)^2}$$

$$= \mathbf{3,211}$$

Lakukan proses perhitungan data diatas sampai sebanyak 31 kali sesuai banyaknya data, hasil perhitungannya bisa dilihat pada tabel 3 didapatkan dari perhitungan *Euclidean Distance* pada iterasi I diatas maka didapatkan perbandingan jarak dari masing-masing *cluster* C<sub>0</sub>, C<sub>1</sub>, C<sub>2</sub>.

d. Mengelompokkan Data Ke Masing-Masing *Cluster*

Setelah data telah diproses dan dihitung pada langkah-langkah diatas maka hasil dari proses diatas dikelompokkan menjadi sebagai berikut :

**Tabel 3** : Pengelompokkan Data Cluster dengan Jarak Terdekat

No	C <sub>0</sub>	C <sub>1</sub>	C <sub>2</sub>	Jarak terdekat C <sub>0</sub>	Jarak terdekat C <sub>1</sub>	Jarak terdekat C <sub>2</sub>
1	0.155	1.181	0.410	1		
2	3.468	4.796	3.211			1
3	3.468	4.796	3.211			1
4	0.100	1.273	0.326	1		
5	0.298	1.042	0.552	1		

6	0.261	1.590	0.000			1
7	0.135	1.424	0.188	1		
8	0.094	1.282	0.316	1		
9	0.383	0.951	0.641	1		
10	3.468	4.796	3.211			1
11	0.880	0.450	1.140		1	
12	0.590	0.758	0.845	1		
13	0.367	0.986	0.617	1		
14	0.157	1.192	0.403	1		
15	0.570	0.827	0.813	1		
16	0.269	1.085	0.516	1		
17	3.468	4.796	3.211			1
18	0.108	1.231	0.361	1		
19	0.702	0.632	0.960		1	
20	0.000	1.330	0.261	1		
21	0.484	0.852	0.741	1		
22	0.505	0.833	0.762	1		
23	0.661	0.671	0.920	1		
24	0.851	0.481	1.110		1	
25	3.468	4.796	3.211			1
26	1.330	0.000	1.590		1	
27	0.732	0.603	0.991		1	
28	0.621	0.711	0.880	1		
29	1.082	0.257	1.341		1	
30	0.854	0.487	1.112		1	
31	3.468	4.796	3.211			1
<b>Jumlah</b>				<b>17</b>	<b>7</b>	<b>7</b>

Bedasarkan proses hitung yang didapatkan pada iterasi I sesuai dengan tabel 4.4 diatas maka telah diperoleh *cluster* dari masing-masingnya dimana hasil dari  $C_0$  berjumlah 17 data, hasil dari  $C_1$  sebanyak 7 data dan hasil dari  $C_2$  sebanyak 7 data.

2) Proses Iterasi II (Kedua)

Dari tabel 4.4 diatas maka didapatkanlah kelompok dari anggota  $C_0$ ,  $C_1$  dan  $C_2$  yang mana hasil kelompok ini untuk mendapatkan nilai *centroid* baru sebagai berikut :

a. Anggota *cluster* 0 ( $C_0$ ) terdiri dari 17 anggota

$$C_1(X_1) = \frac{(0,280+0,320+0,310+0,340+0,320+0,290+0,350+0,350+0,310+0,450 +0,340+0,280+0,240+0,300+0,310+0,270+0,270)}{17}$$

$$C_1(X_1) = \mathbf{0,314}$$

$$C_1(Y_2) = \frac{(0,019+0,019+0,019+0,019+0,019+0,010+0,019+0,018+0,018+0,019 +0,019+0,019+0,019+0,019+0,019+0,019+0,019)}{17}$$

$$C_1(Y_2) = \mathbf{0,018}$$

$$C_1(Z_3) = \frac{(3,610+3,520+3,750+3,370+3,510+3,840+4,040+3,810+3,600+3,990 +3,710+3,560+3,460+3,940+3,960+4,120+4,080)}{17}$$

$$C_1(Z_3) = \mathbf{3,757}$$

b. Anggota *cluster* 1 ( $C_1$ ) terdiri dari 7 anggota

$$C_2(X_1) = \frac{(0,230+0,290+0,270+0,240+0,300+0,300+0,320)}{7}$$

$$C_2(X_1) = \mathbf{0,279}$$

$$C_2(Y_2) = \frac{(0,019+0,020+0,019+0,019+0,019+0,018+0,018)}{7}$$

$$C_2(Y_2) = \mathbf{0,019}$$

$$C_2(Z_3) = \frac{(4,340+4,160+4,310+4,790+4,190+4,540+4,310)}{7}$$

$$C_2(Z_3) = \mathbf{4,377}$$

c. Anggota *cluster* 2 ( $C_2$ ) terdiri dari 7 anggota

$$C_0(X_1) = \frac{(0,0+0,0+0,260+0,0+0,0+0,0+0,0)}{7}$$

$$C_0(X_1) = \mathbf{0,037}$$

$$C_0(Y_2) = \frac{(0,0+0,0+0,019+0,0+0,0+0,0+0,0)}{7}$$

$$C_0(Y_2) = \mathbf{0,003}$$

$$C_0(Z_3) = \frac{(0,0+0,0+3,200+0,9+0,9+0,9+0,9)}{7}$$

$$C_0(Z_3) = \mathbf{0,457}$$

Setelah proses perhitungan diatas selesai, maka akan diperoleh *centroid* baru yang dapat di lihat pada tabel 4.5 dengan nilai sebagai berikut :

**Tabel 4 :** Nilai *Centroid* Baru Iterasi ke II

Centroid	Air	Kotoran	ALB
$C_0$	0,314	0,018	3,757
$C_1$	0,279	0,019	4,377
$C_2$	0,037	0,003	0,457

Setelah didapatkan *centroid* baru selanjutnya dilakukan iterasi kembali seperti proses perhitungan diatas dengan cara menghitung pusat *cluster* 0, *cluster* 1 dan *cluster* 2 kemasing-masing *centroid* menggunakan rumus *Euclidean Distance* (D) pada iterasi II, proses diatas diulangi sampai iterasi tersebut tidak berubah nilai *centroid* pada iterasi sebelum dan sesudahnya, pada penelitian ini berhenti pada iterasi ke 7 setelah nilai *centroid* tidak berubah. Maka hasilnya didapat bahwa *cluster* 0 ( $C_0$ ) untuk kualitas baik terdapat 12 data, *cluster* 1 ( $C_1$ ) untuk kualitas sangat baik terdapat 13 data dan *cluster* 2 ( $C_2$ ) untuk kualitas kurang baik terdapat 6 data. Lebih jelasnya pengelompokan kualitas minyak kelapa sawit dapat dilihat pada tabel 5 sebagai berikut :

**Tabel 5 :** Hasil Pengelompokan Keseluruhan

NO	AIR	KOTORAN	ALB	KEPUTUSAN
1	0.280	0.019	3.610	Baik
2	0.000	0.000	0.000	Kurang Baik
3	0.000	0.000	0.000	Kurang Baik
4	0.320	0.019	3.520	Baik
5	0.310	0.019	3.750	Baik
6	0.260	0.019	3.200	Baik
7	0.340	0.019	3.370	Baik
8	0.320	0.019	3.510	Baik
9	0.290	0.010	3.840	Baik
10	0.000	0.000	0.000	Kurang Baik
11	0.230	0.019	4.340	Sangat Baik
12	0.350	0.019	4.040	Sangat Baik
13	0.350	0.018	3.810	Baik
14	0.310	0.018	3.600	Baik
15	0.450	0.019	3.990	Sangat Baik
16	0.340	0.019	3.710	Baik
17	0.000	0.000	0.000	Kurang Baik
18	0.280	0.019	3.560	Baik
19	0.290	0.020	4.160	Sangat Baik
20	0.240	0.019	3.460	Baik
21	0.300	0.019	3.940	Sangat Baik
22	0.310	0.019	3.960	Sangat Baik
23	0.270	0.019	4.120	Sangat Baik
24	0.270	0.019	4.310	Sangat Baik
25	0.000	0.000	0.000	Kurang Baik
26	0.240	0.019	4.790	Sangat Baik
27	0.300	0.019	4.190	Sangat Baik

28	0.270	0.019	4.080	Sangat Baik
29	0.300	0.018	4.540	Sangat Baik
30	0.320	0.018	4.310	Sangat Baik
31	0.000	0.000	0.000	Kurang Baik

## 5. Kesimpulan

Metode *K-Means Clustering* berhasil diterapkan untuk mengelompokkan kualitas minyak kelapa sawit yang terdiri dari kriteria baik, sangat baik dan kurang baik. Hasil dari penelitian ini dapat mempermudah PT Tasma Puja dalam menentukan kualitas minyak kelapa sawit. Berdasarkan data yang digunakan untuk mengelompokkan kualitas minyak kelapa sawit pada bulan Desember 2017 dengan jumlah 31 data. Mendapatkan hasil pengelompokkan bahwa *cluster* 0 ( $C_0$ ) untuk kualitas baik terdapat 12 data, *cluster* 1 ( $C_1$ ) untuk kualitas sangat baik terdapat 13 data dan *cluster* 2 ( $C_2$ ) untuk kualitas kurang baik terdapat 6 data.

## Referensi

- [1] Krisdiarto, A.W., Sutiarto, Lilik., Widodo, K.H. (2017). “*Optimasi Kualitas Tandan Buah Segar Kelapa Sawit Dalam Proses Panen-Angkut Menggunakan Model Dinamis*”, *AGRITECH*, Vol. 37, No. 1, Februari 2017, ISSN : 0216-0455. <https://doi.org/10.22146/agritech.17015>
- [2] Muzakir, Ari., Wulandari, R.A. (2016). “*Model Data Mining Sebagai Prediksi Penyakit Hipertensi Kehamilan Dengan Teknik Decision Tree*”, *Scientific Journal of Informatics* Vol. 3, No. 1, Mei 2016 ISSN : 2407-7658. <https://doi.org/10.15294/sji.v3i1.4610>
- [3] Abdullah., Usman., Efendi, M. (2017). “*Sistem Klasifikasi Kualitas Kopra Berdasarkan Warna Dan Tekstur Menggunakan Metode Nearest Mean Classifier (NMC)*”, *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)* Vol.4, No.4, Desember 2017, hlm. 297-303 ISSN: 2355-7699. <https://doi.org/10.25126/jtik.201744479>
- [4] Nasari, Fina., Sianturi, C.J.M. (2016). “*Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat*”, *Cogito Smart Journal*/Vol. 2/No. 2/Desember 2016. [https://doi.org/2016:2\(2\):108-119](https://doi.org/2016:2(2):108-119)
- [5] Hakim, Lukman., Seruni, Harvin. (2018). “*Indikasi Penyimpangan Laporan Keuangan Akademik Universitas XYZ Menggunakan Algoritma Greedy dan K-Means*”, *Jurnal Rekayasa Sistem dan Teknologi Informasi*, Vol. 2, No. 1 (2018) 308-313. <https://doi.org/10.29207/resti.v2i1.261>
- [6] Nishom.M.,Fathoni,M.Y.(2018).“*Implementasi Pendekatan Rule-Of-Thumb untuk Optimasi Algoritma K-Means Clustering*”, *Jurnal Pengembangan IT* Vol. 3, No.2. <https://doi.org/10.30591/jpit.v3i2.909>
- [7] Sulastri, Heni., Gufroni, A.I. (2017). “*Penerapan Data Mining Dalam Pengelompokan Penderita Thalassemia*”, *Jurnal Teknologi dan Sistem Informasi - Vol. 03 No. 02 (2017) 299-305*. <https://doi.org/10.25077/TEKNOSI.v3i2.2017.299-305>