

Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



**Penerapan *Data Mining* dalam Pengelompokan Kualitas Produk Kelapa Sawit
Menggunakan Algoritma *K-Means Clustering***

***Application of Data Mining in Classifying the Quality of Palm Products Using
Algorithm K-Means Clustering***

Suan Ekie Nanda Putra¹, Mhd. Furqan²

^{1,2}Prodi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara
Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara
email: ¹ekie.nanda1@gmail.com, ²mfurqan@uinsu.ac.id

ABSTRAK

Minyak kelapa sawit banyak digunakan dalam berbagai produk, termasuk makanan, kosmetik, dan biodiesel. Untuk menjaga kualitas produk, diperlukan pemantauan serta analisis data secara terperinci sangat penting. Pada PT. Sri Ulina Ersada Karina, proses produksi *Crude Palm Oil* saat ini hanya mengikuti standar nasional tanpa analisis lebih lanjut tentang kualitas produk. Dengan analisis yang lebih mendalam, perusahaan dapat meningkatkan efisiensi dan mutu produk. Penelitian ini bertujuan untuk menerapkan teknik *data mining*, khususnya algoritma *K-Means Clustering*, untuk mengelompokkan kualitas produk kelapa sawit yang diolah menggunakan *tools Jupyter Notebook*. Hasil dari penelitian ini menghasilkan 3 *cluster* yaitu *cluster 0* kategori baik dengan jumlah data sebanyak 89 sampel, *Cluster 1* kategori kurang baik dengan jumlah data sebanyak 72 sampel, dan *Cluster 2* kategori sangat baik dengan jumlah data sebanyak 132 sampel.

Kata Kunci: *Data Mining; K-Means Clustering; Crude Palm Oil; Pemrograman Python*

ABSTRACT

The Palm oil is widely used in various products, including food, cosmetics and biodiesel. To maintain product quality, detailed monitoring and data analysis is required. At PT. Sri Ulina Ersada Karina, the current *Crude Palm Oil* production process only follows national standards without further analysis of product quality. With deeper analysis, companies can improve efficiency and product quality. This research aims to apply data mining techniques, especially the *K-Means Clustering* algorithm, to the quality of palm oil products processed using Jupyter Notebook. The results of this research produced 3 clusters, namely cluster 0 in the good category with a total of 89 samples of data, Cluster 1 in the poor category with a total of 72 samples of data, and Cluster 2 in the very good category with the total data is 132 samples.

Keywords: *Data Mining; K-Means Clustering; Crude Palm Oil; Python Programming*

1. PENDAHULUAN

Industri perkebunan kelapa sawit merupakan salah satu sektor ekonomi yang penting di berbagai negara tropis, terutama di Asia Tenggara seperti Indonesia. Kelapa sawit menghasilkan minyak mentah atau *Crude Palm Oil* digunakan dalam berbagai jenis produk, seperti makanan, kosmetik, dan bahan bakar biodiesel. *Crude Palm Oil* (CPO) adalah minyak yang diperoleh dari ekstraksi buah kelapa sawit dan kualitasnya sangat dipengaruhi oleh jenis serta tingkat kematangan buah yang digunakan[2]. Standar nasional untuk CPO di Indonesia mencakup kadar asam lemak bebas, air, dan kotoran yang masing-masing maksimum adalah 5%, 0,25%, dan 0,25%.

Untuk menjaga kualitas dan meningkatkan efisiensi produksi, penting bagi produsen kelapa sawit untuk dapat mengelompokkan produk berdasarkan kualitasnya. PT. Sri Ulina Ersada Karina, sebagai salah satu perusahaan perkebunan kelapa sawit, terus berupaya menjaga dan meningkatkan kualitas produk kelapa sawit yang dihasilkannya. Dalam upaya strategis untuk meningkatkan kualitas produk secara optimal dan efisien, perusahaan perlu memanfaatkan teknologi Data Mining *K-Means Clustering* untuk mengelompokkan kualitas produk kelapa sawit.

Penerapan Data Mining *K-Means Clustering* melibatkan penggunaan algoritma *Clustering* untuk mengelompokkan data berdasarkan kesamaan tertentu[3]. Dalam konteks kualitas CPO, data mining *Clustering* dapat digunakan untuk mengelompokkan produk berdasarkan kadar asam lemak bebas (FFA CPO), kadar air (*Moist* CPO), dan kadar kotoran (*Dirt*). Misalnya, metode *K-Means Clustering* dapat digunakan untuk membagi produk CPO menjadi beberapa kelompok kualitas berdasarkan parameter-parameter tersebut. Dengan demikian, perusahaan dapat dengan cepat mengidentifikasi kelompok produk yang memenuhi standar kualitas, serta kelompok yang memerlukan perhatian khusus untuk perbaikan kualitas.

Dengan penggunaan teknologi ini, PT. Sri Ulina Ersada Karina dapat mengidentifikasi dan memetakan variasi kualitas produk yang dihasilkan, serta mengambil tindakan yang tepat untuk meningkatkan standar kualitas. Pendekatan ini memungkinkan PT. Sri Ulina Ersada Karina untuk mengoptimalkan proses produksi. Data mining *K-Means Clustering* mendukung pengambilan keputusan yang lebih efektif dan efisien [4], memberikan keunggulan kompetitif dalam industri kelapa sawit.

Penelitian tentang pengelompokan data dengan metode *K-Means* dalam bidang data mining telah dilakukan oleh [5] menunjukkan bahwa penerapan metode *K-Means Clustering* mampu mengelompokkan barang-barang lelang ke dalam beberapa *cluster* berdasarkan mutu dan karakteristiknya. Dalam penelitian [6] algoritma *K-Means Clustering* ini dilakukan untuk mengatasi masalah pengelolaan buku yang kurang optimal menyebabkan ketidakseimbangan antara ketersediaan dan permintaan buku. Hasilnya adalah pengelompokan buku menjadi tiga kategori: paling diminati, cukup diminati, dan sedikit diminati.

Penelitian lain yang dilakukan menghasilkan pengelompokan yang menunjukkan terdapat tiga *cluster* yaitu *cluster* 1 (kategori rendah) dengan 12 anggota, *cluster* 2 (kategori cukup) dengan lima anggota, dan *cluster* 3 (kategori tinggi) dengan delapan anggota[7]. Penelitian yang dilakukan oleh [8] menghasilkan penerapan algoritma *K-Means Clustering* dalam sistem informasi klasterisasi menghasilkan klasifikasi pengelompokan data yang efektif.

Berbeda dengan penelitian sebelumnya, penelitian ini akan menerapkan algoritma *k-means clustering* untuk mengelompokkan data produksi *Crude Palm Oil*. Tidak seperti penelitian lain yang memanfaatkan algoritma *K-Means Clustering* untuk berbagai masalah berbeda, seperti klasifikasi jenis obat, jenis buku, wilayah penyebaran COVID-19, dan area produksi tanaman obat. Penelitian ini akan menggunakan data produksi dari PT. Sri Ulina Ersada Karina dan mengelompokkan ke dalam 3 klaster yaitu baik, kurang baik dan sangat baik.

Berdasarkan penjelasan di atas, penelitian ini akan mengelompokkan kualitas produksi kelapa sawit di PT. Sri Ulina Ersada Karina dengan menggunakan data produksi *Crude Palm Oil* dari satu tahun terakhir dan menerapkan algoritma *K-Means Clustering*. Sehingga diharapkan perusahaan dapat dengan mudah mengelompokkan beberapa variasi kualitas produk kelapa sawit secara optimal dan efisien.

1. TINJAUAN TEORI

2.1. Data Mining

Data Mining adalah proses mengekstraksi informasi dan pola yang berharga dari sejumlah besar data[9]. Ini juga sering disebut sebagai penemuan pengetahuan, ekstraksi pengetahuan, analisis data/pola, pengambilan informasi, dan lain sebagainya[10]. Data mining terdiri dari beberapa kategori, yaitu Deskripsi, Klasifikasi, Estimasi, Prediksi, Pengelompokan, dan Asosiasi[11]. Tujuan proses data mining juga disebut sebagai KDD (*knowledge discovery in database*) adalah untuk menggali data dalam kumpulan data dan menghasilkan informasi baru yang bermanfaat[12]. Tujuan data mining adalah untuk menemukan pola dalam kumpulan data, karena itu juga disebut sebagai pengenalan pola.

2.2. Clustering

Clustering atau pengelompokan adalah metode statistik yang digunakan untuk membagi sejumlah besar data berdasarkan karakteristik yang dimiliki oleh objek atau data tersebut[13]. Menurut [14] Konsep *clustering* atau pengelompokan data memiliki karakteristik yang sederhana dan sesuai dengan pola pikir manusia. Saat kita dihadapkan dengan sejumlah besar data, kita cenderung mengelompokkannya menjadi kategori-kategori yang lebih kecil untuk memudahkan analisis lebih lanjut [15]. Selain itu, sebagian besar data yang dikumpulkan dalam berbagai konteks cenderung mempunyai karakteristik yang membentuk pengelompokan alami.

2.3. Algoritma K-Means

Dalam algoritma *K-Means*, data atau objek dikelompokkan ke dalam k kelompok atau klaster[16]. Sebelum melakukan pengelompokan menggunakan algoritma *K-Means*, perlu ditetapkan terlebih dahulu nilai k . Biasanya pengguna sudah memiliki pengetahuan awal mengenai objek yang diteliti, termasuk jumlah klaster yang dianggap paling sesuai. Algoritma *K-Means* menggunakan ukuran *dissimilarity* atau ketidakmiripan untuk mengklaster objek berdasarkan konsep jarak. Objek yang berjarak dekat dianggap mirip[17]. Makin dekat jarak antar objek, makin tinggi tingkat kemiripannya. Sebaliknya, semakin jauh jaraknya makin tinggi pula ketidakmiripannya[18].

3. METODE PENELITIAN

3.1. Metode K-Means Clustering

Dalam penelitian ini proses pembentukan *K-Means Clustering* dilakukan beberapa tahapan, yang mencakup:

- Tentukan k sebagai jumlah cluster yang ingin dibentuk
- Tentukan posisi awal *centroid* (titik pusat cluster) secara random.
- Hitung jarak setiap data ke masing-masing *centroid* menggunakan *euclidean distance*

$$D(ij) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

Keterangan:

$d(ij)$ = Jarak data ke- i ke pusat cluster j

X_{ki} = Data ke-i atribut data ke j

X_{kj} = Data ke-j atribut data ke k

- d. Menentukan *centroid* baru, untuk mendapatkan *centroid* baru dapat menghitung rata-rata dari data-data yang ada pada setiap cluster

$$\mu_i = \frac{1}{N_i} \sum_{q=1}^{N_i} x_q \quad (2)$$

Keterangan:

μ_i = titik *centroid* dari cluster ke-i

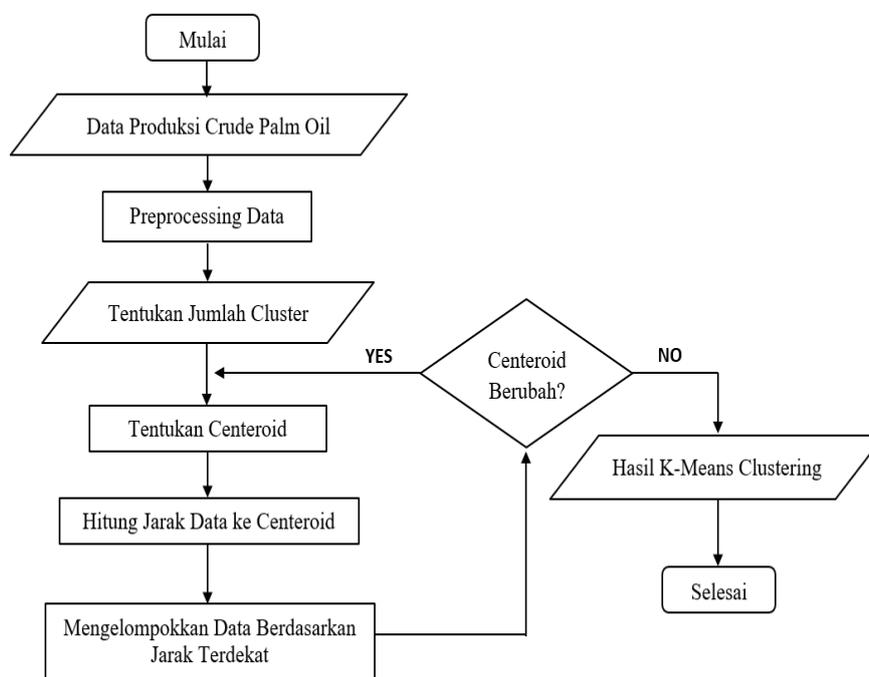
N_i = banyaknya data pada cluster ke-i

x_q = data ke-q pada cluster ke-i

- e. Kembali ke langkah-3 jika *centroid* baru dengan *centroid* lama tidak sama

3.2. Analisis Data Menggunakan K-Means Clustering

Ada beberapa tahapan dalam metode ini, dapat kita lihat dari *flowchart* dibawah ini:



Gambar 1. Flowchart K-Means Clustering

Berdasarkan flowchart diatas maka langkah-langkahnya dapat diuraikan sebagai berikut:

- Input data produksi *Crude Palm Oil* yang mencakup berbagai parameter kualitas seperti FFA CPO, *Moist* CPO, dan *Dirt* CPO.
- Preprocessing* Data. Pada tahap ini mencakup aktivitas seperti normalisasi, transformasi, dan penanganan data yang hilang.
- Tentukan jumlah cluster k
- Tentukan posisi awal *centroid* untuk setiap *cluster*, di mana *centroid* adalah titik tengah dari setiap *cluster* yang akan digunakan sebagai referensi dalam proses *Clustering*.
- Hitung jarak setiap data point terhadap *centroid* menggunakan *Euclidean Distance*.
- Selanjutnya kelompokkan setiap data *point* ke dalam *cluster* dengan jarak terdekat ke *centroid* sehingga data *point* yang paling dekat dengan *centroid* tertentu akan dikelompokkan ke dalam *cluster* tersebut.

- g. Ketika menentukan *centroid*, maka:
 1. Jika *centroid* baru berubah, proses akan diulangi lagi ke langkah-4
 2. Jika *centroid* baru tidak berubah, maka perhitungan selesai

4. HASIL DAN PEMBAHASAN

4.1. Deskriptif Data

Dataset yang akan digunakan pada penelitian ini adalah data produksi *Crude Palm Oil* di PT. Sri Ulina Ersada Karina berdasarkan satu terakhir. Data tersebut terdiri dari 300 data sampel dengan 3 parameter yaitu FFA CPO, *Moist* CPO, dan *Dirt* CPO. Kita dapat melihat data tersebut sebagai berikut:

Tabel 1. Dataset awal *Crude Palm Oil*

No	Tanggal	FFA CPO	Moist CPO	Dirt CPO
1	02-Jan-23	4.99	0.18	0.018
2	05-Jan-23	2.90	0.16	0.015
3	06-Jan-23	2.74	0.17	0.015
4	07-Jan-23	2.62	0.18	0.016
5	10-Jan-23	2.96	0.17	0.015
6	11-Jan-23	2.92	0.16	0.016
7	12-Jan-23	2.95	0.17	0.016
8	13-Jan-23	2.90	0.17	0.015
9	14-Jan-23	2.96	0.18	0.015
10	16-Jan-23	2.99	0.15	0.015
11	17-Jan-23	3.65	0.16	0.016
12	19-Jan-23	3.55	0.17	0.017
13	20-Jan-23	3.16	0.16	0.015
...
...
299	19-Jan-24	2.77	0.17	0.017
300	20-Jan-24	2.79	0.16	0.017

4.2. Preprocessing Data

Dari data yang diperoleh akan dilakukan *preprocessing* data. Ini bertujuan untuk memeriksa kesalahan atau anomali dalam data, mengisi nilai yang hilang (jika ada), dan memastikan konsistensi dalam format data. Pada tahap ini peneliti menyeleksi data dan menghilangkan variabel tanggal karena tidak dibutuhkan dalam proses *Clustering*.

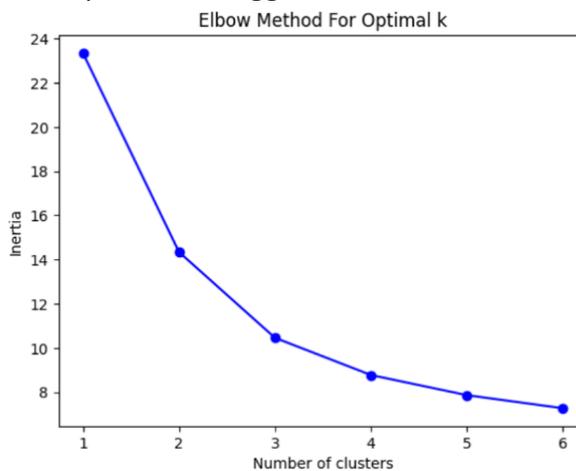
Setelah menyeleksi data, selanjutnya dilakukan proses normalisasi data. Normalisasi dilakukan untuk mengubah skala data sehingga setiap parameter memiliki bobot yang sama dalam analisis *Clustering*. Proses normalisasi data menggunakan pendekatan *Min-Max Scaler* karena dapat menjaga skala asli data dan menjadi solusi dalam penyederhanaan *feature* yang terlalu jauh serta dapat mempertahankan interpretasi data asli pada skala yang relatif. Hasil dari penyeleksian dan normalisasi data ini nantinya yang akan digunakan dalam proses *Clustering*. dapat dilihat pada tabel 2 sebagai berikut.

Tabel 2. Hasil Processing Data

No	FFA CPO	Moist CPO	Dirt CPO
1	0,91475	0,3	0,66667
2	0,22951	0,1	0,16667
3	0,17705	0,2	0,16667
4	0,1377	0,3	0,33333
5	0,24918	0,2	0,16667
6	0,23607	0,1	0,33333
7	0,2459	0,2	0,33333
8	0,22951	0,2	0,16667
9	0,24918	0,3	0,16667
10	0,25902	0	0,16667
11	0,47541	0,1	0,33333
12	0,44262	0,2	0,5
13	0,31475	0,1	0,16667
...
...
299	0,18689	0,2	0,5
300	0,19344	0,1	0,5

4.3. Penerapan K-Means Clustering

Algoritma *K-Means Clustering* diterapkan untuk mengelompokkan data berdasarkan parameter-parameter kualitas *Crude Palm Oil* yang telah dinormalisasi. Salah satu tantangan utama dalam penerapan *K-Means Clustering* adalah menentukan jumlah *cluster* yang optimal. Untuk mengatasi tantangan tersebut, peneliti menggunakan metode *Elbow*.



Gambar 2. Hasil Metode *Elbow*

Dari gambar 2 di atas, terlihat bahwa penurunan grafik mulai melambat setelah $k = 3$. Hal ini menunjukkan bahwa tiga *cluster* merupakan jumlah yang optimal untuk data *Crude Palm Oil* yang telah dinormalisasikan.

Setelah jumlah cluster telah ditentukan, kita akan menentukan *centroid* awal secara acak yang akan digunakan untuk inialisasi *centroid*. Dalam kasus ini, *centroid* pusat awal akan menggunakan C0 dari data ke 3, C1 dari data ke 11 dan C2 dari data ke 5 dari hasil normalisasi data. Tabel berikut menunjukkannya.

Tabel 3. Centroid Awal

Centroid Awal	FFA CPO	Moist CPO	Dirt CPO
Centroid 0	0,17705	0,2	0,16667
Centroid 1	0,47541	0,1	0,33333
Centroid 2	0,24918	0,2	0,16667

Setelah menentukan *centroid* awal, selanjutnya kita akan melakukan perhitungan jarak *centroid* dengan rumus *Euclidean Distance* seperti dibawah ini:

$$D(ij) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2}$$

Iterasi 1

1. Menghitung jarak *centroid* terdekat data ke 3 pada *centroid* 0 berdasarkan nilai atribut (0.17705, 0.2, 0.16667)

$$C0 = \sqrt{(0.17705 - 0.17705)^2 + (0.2 - 0.2)^2 + (0.16667 - 0.16667)^2} = 0$$

$$C1 = \sqrt{(0.17705 - 0.47541)^2 + (0.2 - 0.1)^2 + (0.16667 - 0.33333)^2} = 0.35608$$

$$C2 = \sqrt{(0.17705 - 0.24918)^2 + (0.2 - 0.2)^2 + (0.16667 - 0.16667)^2} = 0.07213$$

2. Menghitung jarak *centroid* terdekat data ke 11 pada *centroid* 1 berdasarkan nilai atribut (0.47541, 0.1, 0.33333)

$$C0 = \sqrt{(0.47541 - 0.17705)^2 + (0.1 - 0.2)^2 + (0.33333 - 0.16667)^2} = 0.35608$$

$$C1 = \sqrt{(0.47541 - 0.47541)^2 + (0.1 - 0.1)^2 + (0.33333 - 0.33333)^2} = 0$$

$$C2 = \sqrt{(0.47541 - 0.24918)^2 + (0.1 - 0.2)^2 + (0.33333 - 0.16667)^2} = 0.29825$$

3. Menghitung jarak *centroid* terdekat data ke 5 pada *centroid* 2 berdasarkan nilai atribut (0.24918, 0.2, 0.16667)

$$C0 = \sqrt{(0.24918 - 0.17705)^2 + (0.2 - 0.2)^2 + (0.16667 - 0.16667)^2} = 0.35608$$

$$C1 = \sqrt{(0.24918 - 0.47541)^2 + (0.2 - 0.1)^2 + (0.16667 - 0.33333)^2} = 0.29825$$

$$C2 = \sqrt{(0.24918 - 0.24918)^2 + (0.2 - 0.2)^2 + (0.16667 - 0.16667)^2} = 0$$

Berdasarkan perhitungan jarak *centroid* terdekat yang telah dilakukan, hasil perhitungan Iterasi 1 dapat dilihat pada tabel di bawah ini.

Tabel 4. Hasil Iterasi 1

Data	Jarak Centroid			cluster
	C0	C1	C2	
1	0,896772708	0,58663037	0,838441068	1
2	0,112924982	0,297056166	0,101916186	2
3	0	0,356081796	0,07213	0
4	0,198302744	0,392489546	0,224061032	0
5	0,07213	0,298254201	0	2
6	0,203122909	0,23934	0,194800995	2
7	0,180321596	0,250349436	0,166692273	2
8	0,05246	0,313436382	0,01967	2
9	0,123299379	0,34489936	0,1	2
10	0,216145971	0,290861114	0,200241918	2
11	0,356081796	0	0,298254201	1

12	0,42618812	0,197114365	0,385393205	1
13	0,170180169	0,231489074	0,119580203	2
...
...
299	0,333475208	0,34788314	0,339100181	0
300	0,348392768	0,327545371	0,352442671	1

Setelah dilakukan perhitungan menggunakan euclidean distance, maka pengelompokan berdasarkan jarak minimum ke *centroid* terdekat adalah

cluster lama (0 0 0 0 0 0 0 0 0 0)

cluster baru (1 2 0 0 2 2 2 2 2 1 1)

Terjadi perubahan pengelompokan data, maka perhitungan iterasi berikutnya akan dilakukan. Sebelum melanjutkan ke iterasi berikutnya, peneliti harus menentukan *centroid* baru dengan mempertimbangkan nilai semua fitur di *cluster 0*, *cluster 1*, dan *cluster 2*.

$$C0 = \left(\frac{0.17705 + 0.1377 + 0.14754 + 0.19672 + 0.2 + \dots + 0.18689}{97} \right) = 0.13429$$

$$= \left(\frac{0.2 + 0.3 + 0.2 + 0.2 + 0.2 + \dots + 0.2}{97} \right) = 0.19072$$

$$= \left(\frac{0.16667 + 0.33333 + 0.16667 + 0.5 + 0.5 + \dots + 0.5}{97} \right) = 0.35190$$

C0 = (0.13429, 0.19072, 0.35190)

$$C1 = \left(\frac{0.91475 + 0.47541 + 0.31475 + 0.44918 + 0.44918 + \dots + 0.19344}{142} \right) = 0.44804$$

$$= \left(\frac{0.3 + 0.1 + 0.2 + 0.2 + 0.3 + \dots + 0.1}{142} \right) = 0.22676$$

$$= \left(\frac{0.66667 + 0.33333 + 0.5 + 0.33333 + 0.33333 + \dots + 0.5}{142} \right) = 0.52230$$

C1 = (0.44804, 0.22676, 0.52230)

$$C2 = \left(\frac{0.22951 + 0.24918 + 0.23607 + 0.2459 + 0.22951 + \dots + 0.22295}{142} \right) = 0.27454$$

$$= \left(\frac{0.1 + 0.2 + 0.1 + 0.2 + 0.2 + \dots + 0.2}{142} \right) = 0.20491$$

$$= \left(\frac{0.16667 + 0.16667 + 0.33333 + 0.33333 + 0.16667 + \dots + 0.33333}{142} \right) = 0.27868$$

C2 = (0.27454, 0.20491, 0.27868)

Berikut adalah tiga *centroid* baru yang akan di iterasi ulang sebagai berikut:

C0 = (0.13429, 0.19072, 0.35190)

C1 = (0.44804, 0.22676, 0.52230)

C2 = (0.27454, 0.20491, 0.27868)

Centroid baru ini yang nantinya akan digunakan dalam perhitungan pada iterasi ke selanjutnya. Setiap iterasi akan menghasilkan *centroid* baru sampai dengan berakhirnya proses iterasi.

Proses iterasi akan terus berlanjut sampai nilai *centroid* dan posisi klaster tidak berubah lagi. Dalam penelitian ini, proses tersebut dilakukan hingga iterasi ke-10. Hasil dari iterasi ke-10 dapat dilihat pada tabel berikut.

Tabel 5. Hasil Iterasi 10

Data	Jarak <i>Centroid</i>			<i>cluster</i>
	C0	C1	C2	
1	0,694920244	0,358218544	0,807163712	1
2	0,411110189	0,523302176	0,148481691	2
3	0,399325793	0,543658412	0,123251621	2

4	0,262319263	0,494870567	0,142521392	2
5	0,395670281	0,489901471	0,123517612	2
6	0,255106993	0,428260388	0,104505376	2
7	0,229206133	0,395804161	0,060317851	2
8	0,395367661	0,504106306	0,119192334	2
9	0,404794974	0,49018512	0,165579589	2
10	0,449841197	0,540993884	0,227490966	2
11	0,351721914	0,259456136	0,281799411	1
12	0,218703736	0,163202038	0,315564305	1
...
299	0,078421243	0,414290203	0,21763116	0
300	0,135208531	0,431291465	0,234325921	0

Pada iterasi ke-10, data yang dikelompokkan menunjukkan hasil yang sama dengan pengelompokan iterasi ke-9, dengan demikian proses pengolahan dataset dihentikan.

4.4. Penerapan K-Means Clustering pada Jupyter Notebook

4.4.1. Import Data

Langkah pertama dalam analisis data menggunakan *Python* adalah mengimport data. *Pandas library*, yang sangat populer dalam analisis data, digunakan untuk mengimport data. Pastikan *Pandas* sudah terinstal di *Jupyter Notebook*. Kemudian, menuliskan baris kode yang cocok dengan *Jupyter Notebook* untuk mengimpor pustaka. Hal ini akan memastikan bahwa *Jupyter Notebook* dapat mengakses perpustakaan dengan benar. Data dapat diproses setelah *import* selesai. Pada tahap *import*, data dataset produksi minyak kelapa sawit crude ini berformat CSV dengan 300 data yang terdiri dari empat atribut yaitu tanggal, FFA CPO, Moist CPO, dan Dirt CPO.

```
[4]: dataInput = pd.read_csv(r'D:\Skripsi\percobaan\Clustering CPO\data_cpo.csv', sep = ";")
      print("Menampilkan Data:")
      print(dataInput)

Menampilkan Data:
   Tanggal  FFA CPO  Moist CPO  Dirt CPO
0  02-Jan-23    4.99    0.18    0.018
1  05-Jan-23    2.90    0.16    0.015
2  06-Jan-23    2.74    0.17    0.015
3  07-Jan-23    2.62    0.18    0.016
4  10-Jan-23    2.96    0.17    0.015
...
295 16-Jan-24    3.70    0.16    0.017
296 17-Jan-24    4.13    0.16    0.017
297 18-Jan-24    2.88    0.17    0.016
298 19-Jan-24    2.77    0.17    0.017
299 20-Jan-24    2.79    0.16    0.017

[300 rows x 4 columns]
```

Gambar 3. Dataset Crude Palm Oil

4.4.2. Preprocessing Data

Preprocessing data adalah tahap penting dalam analisis data yang bertujuan untuk membuat data mentah siap digunakan. Dilakukan penyeleksian dan normalisasi pada data yang akan digunakan. Hasilnya ditampilkan pada gambar dibawah ini.

[11]:

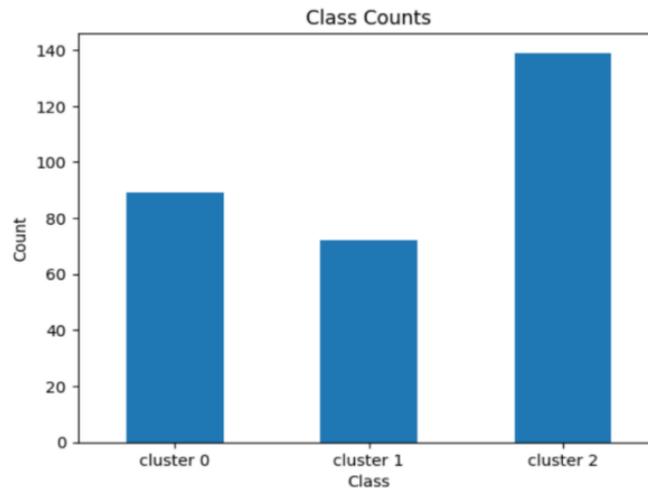
	FFA CPO	Moist CPO	Dirt CPO
0	0.914754	0.3	0.666667
1	0.229508	0.1	0.166667
2	0.177049	0.2	0.166667
3	0.137705	0.3	0.333333
4	0.249180	0.2	0.166667
...
295	0.491803	0.1	0.500000
296	0.632787	0.1	0.500000
297	0.222951	0.2	0.333333
298	0.186885	0.2	0.500000
299	0.193443	0.1	0.500000

300 rows × 3 columns

Gambar 4. Hasil Preprocessing Data

4.4.3. Proses Clustering

Dalam proses *Clustering* dilakukan dengan menggunakan bahasa pemrograman python dengan *tools Jupyter Notebook*. Berdasarkan Gambar 2, jumlah cluster yang optimal untuk data produksi *Crude Palm Oil* adalah $k = 3$. Selanjutnya akan dilakukan proses *Clustering* untuk melihat hasil pengelompokan data. Hasilnya dapat dilihat pada Gambar 5 dan Gambar 6 berikut.



Gambar 5. Hasil Clustering

Class	
cluster 0	89
cluster 1	72
cluster 2	139

Gambar 6. Hasil Pengelompokan K-Means Clustering

Kemudian dilihat juga hasil *centroid* akhir sebagai berikut:

	FFA CPO	Moist CPO	Dirt CPO
0	0.233266	0.213483	0.561798
1	0.598270	0.248611	0.506944
2	0.212667	0.189209	0.284173

Gambar 7. Hasil Centroid Akhir

Berdasarkan perhitungan menggunakan Jupyter Notebook diatas, dapat dilihat bahwa:

1. *cluster 0* kategori baik dengan jumlah data sebanyak 89 sampel. Dapat diketahui bahwa *cluster 0* memiliki karakteristik bebas FFA CPO sedang, *Moist CPO* sedikit lebih tinggi, dan *Dirt CPO* sedang.
2. *cluster 1* kategori kurang baik dengan jumlah data sebanyak 72 sampel. Dapat diketahui bahwa *cluster 1* memiliki karakteristik FFA CPO tinggi, *Moist CPO* sedikit tinggi, dan *Dirt CPO* tinggi.
3. *cluster 2* kategori sangat baik dengan jumlah data sebanyak 132 sampel. Dapat diketahui bahwa *cluster 2* memiliki karakteristik FFA CPO, *Moist CPO*, dan *Dirt CPO* paling rendah, yang menunjukkan kualitas minyak paling tinggi di antara semua klaster.

5. KESIMPULAN

Berdasarkan hasil pengolahan dataset produksi *Crude Palm Oil* di PT. Sri Ulina Ersada Karina menggunakan tools *Jupyter Notebook*, menghasilkan 3 cluter yaitu *cluster 0* kategori baik dengan jumlah data sebanyak 89 sampel. memiliki karakteristik FFA CPO sedang, *Moist CPO* sedikit lebih tinggi, dan *Dirt CPO* sedang. *Cluster 1* kategori kurang baik dengan jumlah data sebanyak 72 sampel. memiliki karakteristik FFA CPO tinggi, *Moist CPO* sedikit tinggi, dan *Dirt CPO* tinggi. *Cluster 2* kategori sangat baik dengan jumlah data sebanyak 132 sampel. memiliki karakteristik FFA CPO, *Moist CPO*, dan *Dirt CPO* paling rendah, yang menunjukkan kualitas minyak paling tinggi di antara semua *cluster*.

REFERENCES

- [1] M. Lambok, F. Sitorus, E. N. Akoeb, R. Sembiring, and M. A. Siregar, "AGRISAINS: Jurnal Ilmiah Magister Agribisnis Peningkatan Produksi Crude Palm Oil Melalui Kriteria Matang Panen Tandan Buah Segar untuk Optimalisasi Pendapatan Perusahaan Improving Crude Palm Oil Production Through Fresh Fruit Harvest Criteria for Optimiza," *J. Ilm. Magister Agribisnis*, vol. 2, no. 1, pp. 26–32, 2020, [Online]. Available: <http://jurnalmahasiswa.uma.ac.id/index.php/agrisains>
- [2] Ainul Haq, Syarifuddin Nasution, and Matri Yanti Hasugian, "Implementasi Plan Do Check Action Pada Produk Crude Palm Oil," *J. Ilm. Tek.*, vol. 3, no. 2, pp. 70–81, 2024, doi: 10.56127/juit.v3i2.1433.
- [3] Haditsah Annur, "Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering (Studi Kasus Toko Luxor Variasi Gorontalo)," *J. Inform. Upgris*, vol. 5, no. 1, pp. 40–45, 2019.
- [4] R. Rokhmatan, N. Suarna, and O. Nurdiawan, "Jurnal Informatika dan Teknologi Informasi Analisa Pengelompokan Dataset Komputer Menggunakan Jurnal Informatika dan Teknologi Informasi," vol. 1, no. 2, pp. 125–132, 2023, doi: 10.56854/jt.v1i2.135.
- [5] M. R. Fahlevi, D. Ridha, D. Putri, and E. Syahrin, "Analisis Pengelompokan Data Pelelangan Barang Dengan Metode K-Means Clustering," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*,

vol. 8, no. 1, pp. 53–61, 2023, [Online]. Available:
<http://tunasbangsa.ac.id/ejurnal/index.php/jurasik/article/view/541>

- [6] N. N. Hasanah and A. S. Purnomo, "Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus : Perpustakaan Politeknik LPP Yogyakarta)," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 4, no. 2, pp. 300–311, 2022, doi: 10.47233/jteksis.v4i2.499.
- [7] N. Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 183–188, 2020, doi: 10.31294/p.v22i2.8808.
- [8] H. A. Siregar, A. Azlan, and N. Y. Lumban Gaol, "Penerapan Data Mining Pada Penjualan Rumah Makan Kasih Ibu Menggunakan Metode K-Means Clustering," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 2, no. 5, p. 750, 2023, doi: 10.53513/jursi.v2i5.8955.
- [9] P. P. Pane, Y. Ramadhan Nasution, and M. Furqan, "Implementasi Data Mining dengan K-Means Clustering untuk Memprediksi Pengadaan Obat," *J. Comput. Syst. Informatics*, vol. 5, no. 2, pp. 286–296, 2024, doi: 10.47065/josyc.v5i2.4920.
- [10] S. Suraya, M. Sholeh, and D. Andayati, "Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa," *Skanika*, vol. 6, no. 1, pp. 51–60, 2023, doi: 10.36080/skanika.v6i1.2982.
- [11] G. Triyandana, L. A. Putri, and Y. Umaidah, "Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means," *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 40–46, 2022, doi: 10.30871/jaic.v6i1.3824.
- [12] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 690–703, 2021, doi: 10.24176/simet.v11i2.5482.
- [13] M. Furqan, Armansyah, and Nurhasanah, "Disease in Corn Leaf Using Gabor Wavelet and K-Means Clustering Algorithm," *J. Mantik*, vol. 5, no. 36, pp. 2152–2156, 2022.
- [14] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.
- [15] D. A. Manalu and G. Gunadi, "Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada Cv Digital Dimensi," *Infotech J. Technol. Inf.*, vol. 8, no. 1, pp. 43–54, 2022, doi: 10.37365/jti.v8i1.131.
- [16] T. Wahyudi, N. Sa, and D. Puspitasari, "Penerapan Metode K-Means Pada Data Penjualan Untuk," vol. 5, no. 1, pp. 228–236, 2023.
- [17] M. Furqan, A. Aulia, U. Islam, N. Sumatera, K. D. Serdang, and S. Utara, "2_4 Kondisi_Ideal_Buah_Jeruk," vol. 6, pp. 199–208, 2022.
- [18] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 186–195, 2019, doi: 10.30812/matrik.v19i1.529.