

IMPLEMENTASI ALGORITMA *K-MEANS CLUSTERING* DALAM MENENTUKAN BLOK TANAMAN SAWIT PRODUKTIF PADA PT ARTA PRIGEL

Yesi Pitaloka Anggriani, Alfis Arif, Febriansyah

Teknik Informatika, Institut Teknologi Pagar Alam

Jl. Masik Siagim No.75, Karang Dalo, Kec. Dempo Tengah, Kota Pagar Alam, Sumatera Selatan 31521

yesipitalokaanggriani@gmail.com

ABSTRAK

Pt Arta Prigel merupakan perusahaan perkebunan kelapa sawit yang telah beroperasi secara komersial sejak tahun 1983 terletak di kota Lahat. Dan memiliki 3 divisi dan 51 blok. Akan tetapi blok blok tanaman sawit tersebut tidak ada rekap bloknya sehingga kurangnya analisis terhadap lahan perkebunan yang mengakibatkan turunnya produksi hasil panen dan salah mengambil keputusan. Tujuan penelitian ini untuk mengimplementasikan metode *K-Means Clustering* dalam menentukan pola hasil produksi sawit yang produktif berdasarkan bloknya di Pt Arta Prigel Lahat. Menggunakan metode pengembangan *CRISP-DM* dan metode pengujian *silhouette coefficient*. Setelah dilakukan proses *clustering* diketahui *cluster_0* dengan tingkat produksi cukup produktif berjumlah 38 blok di tahun 2021 sampai 2023, *cluster_1* dengan tingkat produksi produktif berjumlah 15 blok di tahun 2021 sampai 2023, dan *cluster_2* dengan tingkat tidak produktif berjumlah 47 blok untuk tahun 2021 sampai 2023. Metode pengujian menggunakan *silhouette coefficient* dengan menghitung hasil *silhouette score*. Hasil dari pengujian metode *silhouette coefficient* pada aplikasi *Google Colab* dengan Bahasa Pemrograman *Python* untuk menghitung hasil *silhouette score* terbentuk 3 cluster ($K=3$) dengan nilai 0.61.

Kata kunci : *k-means clustering, Palm, Silhouette Coefficient, Rapid Miner*

1. PENDAHULUAN

Data mining adalah proses mengekstrak informasi yang besar dan kompleks untuk mendapatkan pola tersembunyi dari data yang diproses. Sudah banyak bidang yang menggunakan data mining, salah satunya bidang perkebunan. *Data mining* dapat memberikan gambaran mengenai tingkat produksi dan membantu mengambil keputusan yang tepat. Selain itu, keterbatasan kemampuan manusia dalam melakukan perhitungan dan pengolahan data secara manual tanpa bantuan teknologi membuat perolehan data informasi menjadi terbatas.

Pt Arta Prigel merupakan perusahaan perkebunan kelapa sawit yang telah beroperasi secara komersial sejak tahun 1983. Akan tetapi pada tahun 2017, Pt Arta Prigel memutuskan untuk melakukan penanaman ulang di seluruh areal karena tingkat produksi yang sudah tidak produktif. Selain itu, permasalahan yang muncul pada produktivitas kelapa sawit berasal dari blok-blok tanaman sawit, seperti blok tanaman yang tidak subur, salah dalam memberikan takaran pupuk, kelembapan suhu tanah yang tidak sesuai, dan tidak ada rekap blok yang produktif sehingga kurangnya analisis terhadap lahan perkebunan yang mengakibatkan turunnya produksi hasil panen dan salah dalam mengambil keputusan, dari pemaparan permasalahan yang ada diperlukannya sebuah teknik teknologi yang dapat digunakan untuk menyelesaikan permasalahan secara cepat dan tepat.

Algoritma *k-means clustering* merupakan salah satu algoritma dengan metode analisa data atau proses pemodelan dengan *unsupervised learning*. Algoritma *k-means* merupakan algoritma pengelompokan data yang dapat digunakan untuk membentuk beberapa *cluster* dan dapat diselesaikan tepat waktu. Algoritma

ini berjalan dengan memetakan setiap pengamatan dalam dataset input ke suatu titik dan memilih beberapa titik data awal (k) secara acak, lalu mindahkannya hingga pengelompokan yang paling ideal ditemukan, dapat berupa pola, grafik, dan gambar. Kelebihan dari algoritma *k-means clustering* mampu mengelompokkan data besar dengan cepat.

Clustering adalah metode (*machine learning*) teknik *data mining* untuk menganalisa data dan membentuk kelas objek yang memiliki kesamaan. Dan nantinya akan dimasukkan ke dalam kelompok yang memiliki kesamaan nilai, dengan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster*. Metode *clustering* berguna untuk menemukan kelompok yang tidak dikenal dalam data, memiliki 2 pendekatan dalam mengembangkan metode *clustering* yaitu pendekatan partisi dan hierarki.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining yang dikenal dengan pengenalan pola adalah metode pemrosesan data yang digunakan untuk mendapatkan pola tersembunyi dari data yang akan diproses. Data yang diolah dengan teknik *data mining* akan menciptakan pengetahuan ilmiah baru dari data lama, hasil yang diperoleh dari pengolahan data tersebut dapat digunakan untuk menentukan keputusan di masa depan.[1]

Data Mining atau Penambangan data melibatkan analisis data dari berbagai sudut dan mengubahnya menjadi informasi penting yang dapat digunakan untuk meningkatkan keuntungan. Secara teknis, *data mining* juga dapat digambarkan sebagai proses

menemukan korelasi dan pola di ratusan atau ribuan bidang.[2]

Dari pengertian diatas dapat disimpulkan, *data mining* adalah mengekstrak informasi yang besar dan kompleks untuk mendapatkan pola tersembunyi dari data yang diproses. Penemuan pola ini digunakan untuk menentukan keputusan yang lebih baik di masa mendatang

2.2. Algoritma K-Means

Algoritma *k-means* adalah algoritma dalam pembelajaran tanpa pengawasan (*unsupervised learning*), algoritma ini digunakan untuk mengelompokkan data menjadi beberapa kelompok. *Unsupervised learning* merupakan algoritma *data mining* yang digunakan untuk menemukan pola seluruh atribut menggunakan sistem partisi. Algoritma *k-means* teknik mengelompokkan berbasis jarak yang membagi data menjadi beberapa *cluster* dan hanya bekerja pada angka atau numerik saja.[3]

Algoritma *k-means* merupakan metode *clustering* berbasis jarak yang membagi data menjadi beberapa *cluster*. Ini memberikan kemudahan dan kemampuan untuk menganalisa data besar dengan cepat dan dapat menemukan *outlier*. *K-means* memiliki konsep dimana K elemen data dipilih sebagai pusat kemudian dihitung jarak seluruh elemen data menggunakan rumus *Euclidean*. [4]

Dari penjelasan diatas algoritma *k-means* pengelompokkan data menjadi beberapa *cluster* dapat diselesaikan tepat waktu. Algoritma *k-means* memberikan kemudahan untuk menganalisa data besar dengan cepat dan tepat

2.3. Clustering

Clustering mengacu pada pengelompokkan seperti catatan, observasi atau perhatian dan membentuk kelas-kelas objek yang memiliki kemiripan. *Cluster* adalah sekumpulan *record* yang serupa dan berbeda dengan *record* di cluster lain. *Clustering* berupaya untuk membagi seluruh kumpulan data ke dalam kelompok-kelompok yang relatif sama, dimana kesamaan *record* pada satu kelompok akan mempunyai nilai maksimal, sedangkan kesamaan *record* pada kelompok lain akan mempunyai nilai minimum.[3]

Clustering adalah metode pembelajaran mesin (*machine learning*) yang digunakan untuk mengelompokkan data ke dalam kelompok atau *cluster* yang sesuai. Tujuan dari *clustering* untuk menemukan struktur pada data yang sebelumnya tidak diketahui. *Clustering* merupakan suatu metode yang digunakan untuk mencari dan mengelompokkan data yang mempunyai karakteristik serupa antara data satu dengan data lainnya. [5]

Dari penjelasan diatas disimpulkan bahwa *clustering* adalah teknik *data mining* untuk menganalisa data dan membentuk kelas objek yang memiliki kesamaan. Dan nantinya akan dimasukkan ke dalam kelompok yang memiliki kesamaan nilai.

2.4. Rapid Miner

Rapid miner adalah *platform* perangkat lunak untuk ilmu data dan pembelajaran mesin. *Rapid miner* menyediakan banyak alat untuk persiapan data, pemodelan, evaluasi, dan implementasi. *Rapid miner* dirancang untuk kemudahan penggunaan dan memungkinkan pengguna membuat dan menguji berbagai model dengan mudah, bahkan tanpa pengalaman pemrograman. [5]

Rapid miner menggunakan metode berorientasi objek dalam hierarki *Java* dan dapat digunakan di hampir semua *platform*. Sistem operasi *rapid miner* mencakup alat untuk pemrosesan entitas, klasifikasi, pengelompokkan, regresi, kolaborasi, dan visualisasi. [6]

Dari penjelasan diatas dapat disimpulkan *Rapid miner* sebuah perangkat lunak yang digunakan untuk menguji berbagai model dengan mudah. *Rapid miner* mengevaluasi pengklasifikasi pemrosesannya.

2.5. Python

Python bahasa pemrograman tingkat tinggi yang dapat ditafsirkan, *multiplatform* (berfungsi pada berbagai platform termasuk *MS Windows*, *Linux*, dan *Macintosh*). *Python* dikembangkan oleh Guido Van Rossum di Amsterdam pada tahun 1990 pengembangan lebih lanjut dari bahasa pemrograman ABC. *Python* gratis untuk digunakan siapa saja. [7]

Python merupakan bahasa pemrograman yang ditafsirkan untuk tujuan umum. Berbeda dengan bahasa pemrograman lainnya, *Python* menekankan keterbacaan kode agar sintaksnya lebih mudah dipahami. Hal ini memudahkan untuk mempelajarinya baik seorang pemula atau seseorang yang ahli dalam bahasa pemrograman. [8]

Dari penjelasan diatas disimpulkan bahwa *python* merupakan bahasa pemrograman tingkat tinggi yang mudah digunakan oleh penggunanya, bersifat *interpretatif dan multiplatform* (bisa diinstal di berbagai platform).

2.6. Silhouette Coefficient

Silhouette Coefficient merupakan metode evaluasi atau validasi algoritma *clustering* yang mengukur kualitas *cluster* yang terbentuk. *Silhouette Coefficient* menunjukkan seberapa baik data ditempatkan dalam *cluster* tertentu. [9]

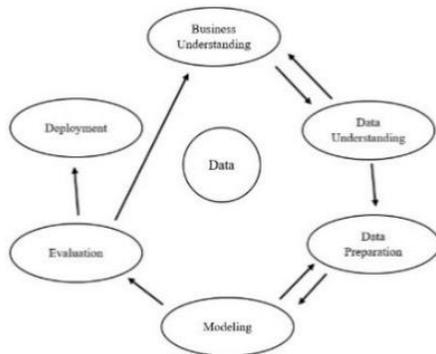
Silhouette coefficient merupakan metode pengujian kualitas *cluster* yang digunakan untuk mengetahui tingkat kepemilikan setiap objek dalam suatu *cluster*. *Silhouette coefficient* juga merupakan salah satu metode integrasi dalam validasi *clustering*, yaitu metode penggabungan yang digunakan untuk menentukan hubungan antar objek dalam *cluster*, dan metode pemisahan yang digunakan untuk menentukan jarak antar *cluster*. [10]

Validasi *Silhouette Coefficient* peneliti dapat mengukur sudah berapa baik *cluster* yang sudah

terbentuk ditempatkan pada *clusternya*, dengan mengukur jarak antara *clusternya* dan *cluster* lain

3. METODE PENELITIAN

Penelitian ini menggunakan teknik data mining yaitu k-means clustering. Metode pengembangan dengan CRISP-DM dengan 6 tahapan yaitu business understanding, data understanding, data preparation, modelling, evaluation, dan deployment. Metode pengujian silhouette coefficient.



Gambar 1. Crisp-Dm

- 1) **Business Understanding**
 Pada tahapan ini ada beberapa yang harus dilakukan yaitu, memahami kebutuhan dan tujuan dari sudut pandang bisnis, lalu mengartikan pengetahuan ke dalam bentuk pendefinisian masalah pada *data mining*, kemudian menentukan rencana serta strategi untuk mencapai tujuan *data mining*.
- 2) **Data Understanding**
 Pada tahapan ini, langkah pertama yang harus dilakukan yaitu pengumpulan data, lalu mendeskripsikan data, dan mengevaluasi kualitas data.
- 3) **Data Preparation**
 Pada tahapan ini, membangun data set dari data mentah. Beberapa hal yang harus dilakukan yaitu, pembersihan data (*Data Cleaning*), pemilihan data (*Data Selection*), record, atribut-atribut, dan transformasi data (*Data Transformation*) yang akan dijadikan masukan dalam tahanan pemodelan.
- 4) **Modeling**
 Pada tahapan ini, akan melibatkan *Machine Learning* langsung dalam penentuan teknik, alat bantu, dan algoritma *data mining*. Pemodelan yang akan digunakan pada penelitian ini adalah metode *clustering* dengan menggunakan algoritma *k-means*.
 - Siapakan dataset
 - Menentukan Jumlah Cluster
 - Titik Pusat cluster
 - Hitung Jarak data
 - Pusat Cluster Berubah

- Mengulang langkah 3 sampai 5
- 5) **Evaluastion**
 Pada tahapan ini, melihat tingkat performa dari pola yang dihasilkan oleh algoritma. berfokus pada model yang dihasilkan sudah sesuai standar *K Means Clustering* dan tidak melewati satu pun tahapan sampai selesai. Tahapan pengujian dilakukan menggunakan metode *Silhouette Coefficient*.
 - 6) **Deployment**
 Pada tahapan ini, melakukan pembuatan laporan dan artikel jurnal menggunakan model yang dihasilkan.

3.1. Pengumpulan Data

Tahap pertama pada penelitian ini yaitu pengumpulan data. Data produksi selama 2 tahun mulai bulan Oktober 2021 sampai Oktober 2023 yang nantinya akan diolah untuk dijadikan dataset clustering hasil produksi. Data Tabel produksi dapat dilihat pada tabel 1.

Tabel 1. Dataset Tahun 2021-2023

No	Blok	Ha	Hasil	Bulan	Tahun
1	1	21.09	5.772	10	2021
2	1	21.09	5.502	11	2021
3	1	21.09	9.120	12	2021
4	2	37.40	5.694	10	2021
5	2	37.40	7.921	11	2021
6	2	37.40	16.413	12	2021
7	13	39.97	8.899	10	2021
8	13	39.97	8.898	11	2021
9	13	39.97	11.131	12	2021
10	14	46.07	7.752	10	2021
.....
1015	51	20.12	17.322	10	2023

4. HASIL DAN PEMBAHASAN

4.1. Business Understanding

Pada pemahaman bisnis dilakukan proses menentukan tujuan dan lingkungan pencarian. Tujuan dari penelitian ini untuk Mengelompokkan hasil produksi berdasarkan blok, dan mencari cluster tertinggi. Maka dari itu peneliti akan menclustering hasil produksi tanaman sawit menggunakan metode *k-means* pada Pt Arta Prigel untuk mengetahui pola hasil produksi.

4.2. Data Understanding

Pada proses pemahaman data ini, data didapat dari bagian KrEstate Pt Arta Prigel yaitu data hasil produksi selama 2 tahun dimulai pada bulan Oktober tahun 2021 sampai dengan bulan Oktober 2023 dengan jumlah 1.275 *record* dengan 5 atribut yaitu Blok, Hektar, Hasil, Bulan, dan Tahun kategori data yang diterima dalam bentuk *excel* dan data yang sudah didapatkan perlu dilakukan proses *cleaning* dan akan dilakukan pemilihan data dengan atributnya.

4.3. Data Preparation

Pengolahan data dilakukan dengan langsung mengimplementasikan pada *Rapid Miner* dengan melakukan tiga tahapan yang sudah ditetapkan yaitu:

4.3.1. Data Selection

Pada tahapan ini dari 5 atribut yang didapat dari bagian KrEstate Pt Arta Prigel diantaranya Divisi, Blok, Luas Lahan, Produksi Perbulan dan Total. Dengan total 51 blok. dan hasil *record* produksi 1.275 data setelah di *cleaning* menjadi 1.015 data yang akan dijadikan *dataset* dengan 5 atribut baru yaitu Blok, Hektar, Hasil, Bulan, dan Tahun. Selanjutnya peneliti melakukan analisa hasil produksi dari setiap blok menggunakan *excel*. Kemudian dataset dari hasil tersebut dapat dilihat pada gambar 2.

	A	B	C	D	E
1	BLOK	HEKTAR	HASIL	BULAN	TAHUN
2	1.000	21.69	4779.000	10.000	2021.000
3	1.000	21.69	5000.000	10.000	2021.000
4	1.000	21.69	6123.000	10.000	2021.000
5	2.000	37.40	6991.000	10.000	2021.000
6	2.000	37.40	7021.000	10.000	2021.000
7	2.000	37.40	10441.000	10.000	2021.000
8	13.000	39.67	8899.000	10.000	2021.000
9	13.000	39.67	8899.000	10.000	2021.000
10	13.000	39.67	11131.000	10.000	2021.000
11	14.000	46.67	7752.000	10.000	2021.000
12	14.000	46.67	11372.000	10.000	2021.000
13	14.000	46.67	11099.000	10.000	2021.000
14	15.000	32.64	8553.000	10.000	2021.000
15	15.000	32.64	10428.000	10.000	2021.000
16	15.000	32.64	12419.000	10.000	2021.000
17	16.000	33.71	1312.000	10.000	2021.000
18	16.000	33.71	1618.000	10.000	2021.000
19	16.000	33.71	1348.000	10.000	2021.000
20	17.000	39.60	11504.000	10.000	2021.000

Gambar 2. Data Selection

4.3.2. Data Processing

Proses *processing* merupakan proses yang mencakup *cleaning* dan *transformasi* data. Pada tahap ini, sudah dipastikan tidak ada lagi data yang kosong sehingga dapat dilihat pada gambar 3.

Name	Type	Min	Max	Mean
ID	Nominal	0	51	1012
BLOK	Nominal	0	51	29.299
HEKTAR	Real	0	11.490	33.617
HASIL	Nominal	0	11490	74826
BULAN	Nominal	0	12	6.239
TAHUN	Nominal	0	2023	2022.544

Gambar 3. Data Processing

4.3.3. Data Transformation

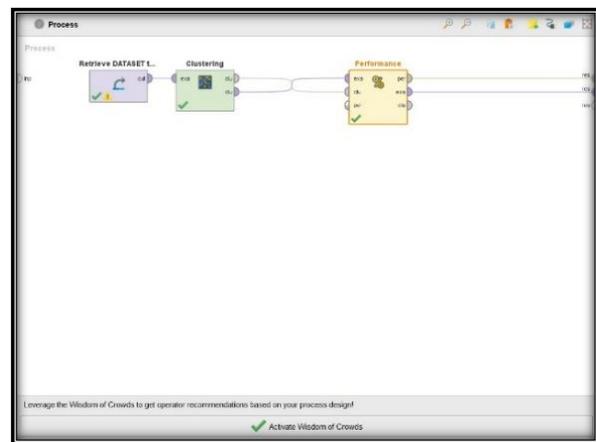
Coding adalah proses mengubah data terpilih sehingga data tersebut cocok untuk proses data mining. Pada proses transformasi data, data yang diproses menunjukkan atribut yang dipilih dan diproses di *Rapid Miner*. Data hasil produksi akan di transformasi berdasarkan kriteria-kriteria yang telah ditentukan. Berikut kriteria produksinya dapat dilihat pada tabel 2

Tabel 2. Kriteria Nilai

Kriteria	Nilai
Cukup Produktif	1
Produktif	2
Tidak Produktif	3

4.3.4. Modeling

Pada proses ini melakukan proses dimana model digunakan dengan algoritma yang digunakan dalam metode *clustering* yaitu operator *clustering* model *k-means* mengambil objek dari port input dan mengirimkan salinannya ke output. Setiap port yang terhubung membuat salinan tidak terikat (*independent*), setiap mengubah satu salinan tidak mempengaruhi salinan yang lain sehingga dapat dihubungkan dengan *performance* yaitu untuk mengetahui kualitas suatu model algoritma *k-means*. Menghasilkan pola informasi yang dapat memudahkan pihak yang berkepentingan seperti pada gambar 4.



Gambar 4. Model Algoritma K-Means

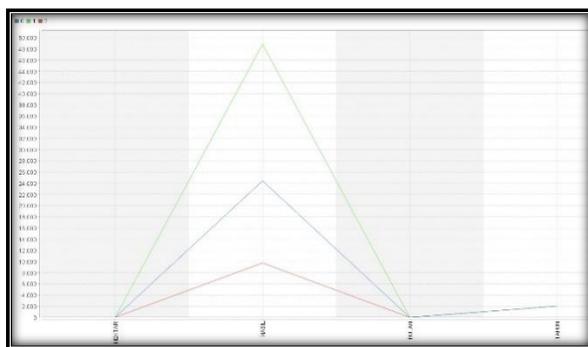
4.3.5. Tentukan Jumlah Cluster

Untuk menentukan jumlah cluster perlu dilakukan percobaan dengan jumlah 3 cluster, yaitu C0, C1, dan C2 dengan atribut Blok, Hektar, Bulan dan tahun 2021 sampai 2023. Dengan pengukuran *performance vector* rata-rata dalam *centroid distance* 23000739.091, rata-rata *cluster_0* 28064575.850, rata-rata *cluster_1* 89406062.799, dan rata-rata dalam *centroid cluster_2* 13922135.927 dan nilai *Davies Bouldin Index* adalah 0.511. dapat dilihat pada gambar 5.

Attribute	cluster_0	cluster_1	cluster_2
HEKTAR	32.885	34.750	30.923
HASIL	24416.787	48861.707	6734.185
BULAN	6.882	6.931	6.306
TAHUN	2022.579	2022.569	2022.192

Gambar 5. Percobaan 3 cluster

1. Tentukan Titik Centroid Secara acak
Titik *centroid* ditentukan secara acak berdasarkan beberapa eksperimen klaster di Rapid Miner. Dari ketiga pengujian tersebut dipilih titik pusat terkecil untuk mendapatkan hasil terbaik.
2. Hitung Jarak data Ke Centroid
Titik *centroid* ditentukan secara acak berdasarkan beberapa eksperimen klaster di Rapid Miner. Dari ketiga pengujian tersebut dipilih titik pusat terkecil untuk mendapatkan hasil terbaik. Pada tahap ini, menghasilkan nilai jarak data *cluster_0* dimulai 10.00 sampai 24.00 untuk *cluster* cukup produktif, *cluster_1* dimulai 24.00 sampai 48.00 untuk *cluster* produktif, dan *cluster_2* dimulai 10.00 sampai 0 untuk *cluster* tidak produktif. Dapat dilihat pada gambar 6.



Gambar 6. Jarak Data ke Centroid

3. Perbarui Nilai Centroid
Setelah beberapa percobaan, centroid pada 3 cluster tidak berubah. Jika titik berat berubah maka iterasi diulangi, tetapi jika tidak berubah maka pengulangan dihentikan dan diperoleh hasil dari setiap kelompoknya.

Berdasarkan perhitungan *Rapid Miner* maka diperoleh pola yang nantinya akan diimplementasikan dalam *Python*. Pola yang digunakan untuk mengelompokkan *Cluster* data berdasarkan hasil perhitungan jarak adalah :

- a. Jika $C_0 < C_1$ dan $C_0 < C_2$ maka *Cluster_0* dengan keterangan Cukup Produktif
- b. Jika $C_1 < C_0$ dan $C_1 < C_2$ maka *Cluster_1* dengan keterangan Produktif
- c. Jika $C_1 < C_0$ dan $C_2 < C_1$ maka *Cluster_2* dengan keterangan Tidak Produktif

Dari data produksi yang diolah dengan *Rapid Miner* yang digunakan menggunakan metode *Clustering K-Means* maka diperoleh pola hasil produksi perblok pada tahun 2021-2023 yaitu *Cluster_0* memiliki tingkat produksi cukup produktif dengan Jumlah 342 data, *Cluster_1* memiliki tingkat produksi produktif dengan jumlah 58 data, dan *Cluster_2* memiliki tingkat produksi Tidak produktif dengan jumlah 615 data

4. Evaluation
Pengujian dilakukan menggunakan *Silhouette Coefficient* terhadap cluster yang sudah terbentuk dengan seluruh data 1015 record. Kemudian didapatkan hasil terbaik dari silhouette score setelah dilakukan beberapa kali percobaan, maka didapatkan hasil *cluster* yang tepat berjumlah 3 *cluster* ($K=3$) dengan nilai kualitasnya yaitu 0,61. Hasil *silhouette* dapat dilihat pada gambar 7.

```

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:878: FutureWarning: The default value of 'n_init'
warnings.warn(
For n_clusters = 2, the average silhouette_score is: 0.624755188160137
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:878: FutureWarning: The default value of 'n_init'
warnings.warn(
For n_clusters = 3, the average silhouette_score is: 0.610567128380889
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:878: FutureWarning: The default value of 'n_init'
warnings.warn(
For n_clusters = 4, the average silhouette_score is: 0.580794729038101
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:878: FutureWarning: The default value of 'n_init'
warnings.warn(
For n_clusters = 5, the average silhouette_score is: 0.553998578846412
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:878: FutureWarning: The default value of 'n_init'
warnings.warn(
For n_clusters = 6, the average silhouette_score is: 0.525797464254009
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:878: FutureWarning: The default value of 'n_init'
warnings.warn(
For n_clusters = 7, the average silhouette_score is: 0.536658788138737
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:878: FutureWarning: The default value of 'n_init'
warnings.warn(
For n_clusters = 8, the average silhouette_score is: 0.543688951568519
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:878: FutureWarning: The default value of 'n_init'
warnings.warn(
For n_clusters = 9, the average silhouette_score is: 0.5485963045857651
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:878: FutureWarning: The default value of 'n_init'
warnings.warn(

```

Gambar 7. Nilai Silhouette Score

5. Deployment
Pada tahap *deployment* merupakan tahapan terakhir berupa pengetahuan atau informasi yaitu mengenai pola hasil produksi Pt Arta Prigel selama 2 tahun sehingga dapat diketahui ada 3 *cluster*, yang pertama *cluster_0* adalah cukup produktif dengan total 38 blok pada tahun 2021 sampai 2023, dimana , kemudian *cluster_1* memiliki keterangan produktif dengan total 15 blok pada tahun 2021 sampai 2023, dan *cluster_2* dengan keterangan Tidak produktif dengan total 47 blok di tahun 2021 sampai 2023. Dari proses *clustering* dapat dilihat tingkat produksi produktif memiliki peningkatan pada tahun 2022, dan untuk hasil produksi pada Pt Arta Prigel lebih dominan berada pada tingkat produksi tidak produktif.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang sudah dilakukan menghasilkan Clustering hasil produksi Pada Pt Arta Prigel. Hasil proses clustering data dalam menentukan blok produktif terbentuk menjadi 3 cluster yaitu *cluster_0* dengan keterangan Cukup Produktif, *cluster_1* dengan keterangan Produktif, dan *cluster_2* dengan keterangan Tidak Produktif.. Diketahui produksi untuk *cluster_0* berjumlah 38 blok, *cluster_1* berjumlah 15 blok, dan *cluster_2* berjumlah 47 blok. Kemudian hasil pada tahap pengujian menggunakan *Silhouette Coefficient* pada aplikasi *google colab* dengan bahasa pemrograman *python* untuk menghitung hasil dari *silhouette score* diperoleh jumlah *cluster* yang tepat yaitu $K=3$ dengan jumlah nilai 0.61. Dari hasil tersebut akan didapatkan beberapa pengetahuan yang diharapkan dapat bermanfaat bagi pihak perusahaan Pt Arta Prigel terutama bagian Manager dan Asisten Kebun untuk penunjang keputusan dan melakukan tindakan selanjutnya.

Perlu adanya analisis yang tepat pada tanaman sawit di setiap bloknnya untuk melihat hasil produksi yang tepat, Perlu dilakukan proses penggalian variasi variabel dan atribut data agar hasil *clustering* yang dihasilkan dapat lebih maksimal. Percobaan dengan algoritma lain dapat digunakan sebagai perbandingan agar dapat menemukan algoritma yang benar-benar cocok untuk mengetahui pola hasil produksi .

DAFTAR PUSTAKA

- [1] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," pp. 762–767, 2019.
- [2] R. I. O. Limabri, F. Putrawansyah, and A. Arif, "Penerapan Data Mining Untuk Mengklasifikasi Nasabah Bank Sumsel Babel Menggunakan Algoritma C4. 5," *Escaf*, pp. 1101–1108, 2023.
- [3] & Z. 2021 Zulfa, Auliya, Permata, "Analisis Data Mining Untuk Clustering Kasus COVID-19 di Provinsi Lampung Dengan Algoritma K-Means," vol. 2, no. 2, pp. 100–108, 2021.
- [4] M. A. K-means, A. Lili, and S. Widodo, "Pengelompokan Hasil Panen Kelapa Sawit Dalam Produksi Per Blok," vol. 01, no. 01, pp. 45–54, 2022.
- [5] M. R. Nahjan, N. Heryana, A. Voutama, F. I. Komputer, U. S. Karawang, and R. Miner, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell," vol. 7, no. 1, pp. 101–104, 2023.
- [6] S. Rokhanah, A. Hermawan, and D. Avianto, "Pengaruh Principal Component Analysis Pada Naïve Bayes dan K-Nearest Neighbor Untuk Prediksi Dini Diabetes Melitus Menggunakan Rapidminer," vol. 11, no. 1, 2023.
- [7] K. Ademariana, F. R. Lumbanraja, and R. Andrian, "Jurnal Pepadun Clustering K-Means Jenis Kata Pada Laporan Kegiatan Kuliah Kerja Nyata (KKN) Universitas Lampung Menggunakan Word2vec © 2021 Ilmu Komputer Unila Publishing Network all rights reserved Jurnal Pepadun," vol. 3, no. 2, pp. 221–228, 2021.
- [8] A. Ridho and C. R. Niani, "Implementasi Enkripsi Dengan Vigenere Cipher Dan Reverse Cipher Menggunakan Bahasa Pemrograman Python," pp. 9–15, 2022.
- [9] M. Astiningrum, M. Mentari, and Y. G. Maranatha, "Mutu Buah Salak Menggunakan Pengolahan Citra Digital," pp. 205–210.
- [10] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.