

## Pengelompokkan Hasil Produksi Tanaman Perkebunan Berdasarkan Provinsi Menggunakan Metode K-Means

### *Clustering Production of Plantation Crops by Province Using the K-Means Method*

Azhari Abdillah Simangunsong<sup>1</sup>, Indra Gunawan<sup>2</sup>, Zulaini Masruro Nasution<sup>3</sup>  
STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

#### Article Info

##### **Genesis Artikel:**

Diterima, 18 Oktober 2022  
Direvisi, 28 Desember 2022  
Disetujui, 25 Januari 2023

##### **Kata Kunci:**

Hasil Produksi  
K-Means  
Pengelompokan  
Provinsi  
Tanaman Perkebunan

#### ABSTRAK

Tujuan dari penelitian ini untuk melakukan pengelompokan hasil produksi tanaman perkebunan setiap tahunnya berdasarkan provinsi di indonesia, agar dapat diketahui provinsi mana saja yang menghasilkan paling banyak produksi tanaman perkebunan maupun yang kurang menghasilkan. Pada penelitian ini menggunakan teknik Data Mining Algoritma K-Means. Sumber data penelitian ini dikumpulkan berdasarkan data perkebunan yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia. Data yang digunakan adalah data dari tahun 2018-2020 yang terdiri dari 34 provinsi. Hasil penelitian ini berupa pengelompokan yang terbagi kedalam 3 Cluster yaitu Cluster rendah, Cluster sedang, dan Cluster tinggi. Berdasarkan hasil perhitungan menggunakan Algoritma K-Means diperoleh Cluster tinggi sebanyak 6 items (Provinsi), Cluster sedang sebanyak 2 Provinsi dan Cluster rendah sebanyak 27 Provinsi. Kesimpulan yang dapat diperoleh, bahwa pengelompokan produksi tanaman perkebunan di Indonesia dapat diselesaikan dengan menerapkan algoritma K-Means.

#### ABSTRACT

The purpose of this research is to classify the results of plantation crop production each year based on provinces in Indonesia, so that it can be known which provinces produce the most plantation crop production and which produce less. In this study using the K-Means Algorithm Data Mining technique. The data source for this research was collected based on plantation data obtained from the Indonesian Central Bureau of Statistics (BPS). The data used is data from 2018-2020 which consists of 34 provinces. The results of this study are groupings which are divided into 3 Clusters, namely low Clusters, medium Clusters, and high Clusters. Based on the results of calculations using the K-Means Algorithm, 6 items (Provinces) were obtained for high Clusters, 2 Provinces for medium Clusters and 27 Provinces for low Clusters. The conclusion that can be obtained is that the grouping of plantation crop production in Indonesia can be solved by applying the K-Means algorithm.

This is an open access article under the [CC BY-SA license](#).



#### **Penulis Korespondensi:**

Azhari Abdillah Simangunsong,  
Program Studi Teknik Informatika,  
STIKOM Tunas Bangsa, Pematangsiantar, Indonesia  
Email: azharisimangunsong@gmail.com

## 1. PENDAHULUAN

Perusahaan perkebunan adalah suatu perusahaan yang berbentuk badan usaha/badan hukum yang bergerak dalam kegiatan budidaya tanaman perkebunan diatas lahan yang dikuasai, dengan tujuan ekonomi/komersil dan mendapat izin usaha dari instansi yang berwewenang dalam pemberian izin usaha perkebunan. Produksi kebun atau lazim disebut produksi primer adalah produksi/hasil yang dipanen dari usaha perkebunan nya tanpa melalui proses pengolahan lebih lanjut. Contoh produksi kebun adalah karet, kelapa sawit, kelapa dan masih banyak lagi.

Indonesia sebagai salah satu negara yang banyak memiliki daerah perkebunan yang memproduksi tanaman berupa kelapa sawit, kelapa dan karet. Tiga produksi tanaman ini merupakan tanaman perkebunan sebagai devisa di Indonesia . Setiap provinsi memiliki keunggulan hasil produksi tanaman yang disesuaikan dengan keadaan tanah dan lahan yang dimiliki provinsi tersebut. Berdasarkan data BPS tentang hasil produksi tanaman perkebunan belum memiliki pengelompokan data berdasarkan hasil produksi tanaman perkebunan provinsi mana yang memiliki produksi tanaman perkebunan seperti sawit, kelapa dan karet terbanyak, sedang dan rendah.

Berdasarkan data tersebut maka perlu dilakukan penelitian dengan menerapkan salah satu algoritma dari cabang ilmu komputer. Sebagaimana diketahui banyak penelitian-penelitian berbasis komputerisasi telah mampu memecahkan banyak masalah yang sifatnya statistik, kelompok, rumit dan saling berkaitan. Diantaranya seperti untuk pendukung keputusan [1]–[8], untuk prediksi dan peramalan menggunakan sistem kecerdasan buatan [9]–[18], hingga di bidang penambangan data khususnya klasifikasi dan *Clustering* [19]–[27]. algoritma *K-Means* untuk mengelompokkan hasil produksi tanaman perkebunan berdasarkan provinsi dengan tiga *Cluster* yaitu *Cluster* terbanyak, *Cluster* sedang, dan *Cluster* terendah untuk tanaman sawit, kelapa dan karet di Indonesia. Alasan menggunakan algoritma *K-Means* diantaranya adalah karena algoritma ini memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar. Selain itu *K-Means* ini tidak terpengaruh terhadap urutan objek.

Algoritma *K-Means* penggunaannya sudah banyak di terapkan. Beberapa penelitian yang sudah pernah dilakukan seperti pada penelitian (Haryadi, 2021) yang menerapkan algoritma K-Means untuk melakukan *Clustering* produksi perkebunan kelapa sawit menurut provinsi. Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa penerapan Algoritma *K-Means* telah berhasil dijadikan solusi dalam menyelesaikan masalah dengan menunjukkan sebuah wawasan baru yaitu pengelompokan provinsi-provinsi yang menjadi prioritas terhadap produksi perkebunan kelapa sawit di indonesia berdasarkan 3 *Cluster*. *Cluster* 1 merupakan kategori provinsi dengan produksi perkebunan kelapa sawit rendah atau Low yaitu 14 dari 21 kategori provinsi yang diuji, kemudian *Cluster* 2 adalah kategori provinsi dengan produksi perkebunan kelapa sawit sedang atau Medium yaitu 4 dari 21 kategori provinsi yang diuji, dan terakhir adalah *Cluster* 3 merupakan kategori provinsi dengan produksi perkebunan kelapa sawit tinggi atau High yaitu 3 dari 21 kategori provinsi yang diuji [28]. Penelitian berikutnya dilakukan oleh (Lili & Widodo, 2022) untuk melakukan pengelompokan hasil panen kelapa sawit berdasarkan produksi per blok dengan menerapkan algoritma *K-Means*". Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa hasil yang diperoleh dari metode *K-means* yang di implementasikan ke dalam *Rapidminer* memiliki nilai yang sama yaitu menghasilkan 3 *Cluster* yaitu *Cluster* tinggi, *Cluster* sedang, dan *Cluster* rendah. Dengan *Cluster* tinggi memiliki 19 blok, *Cluster* sedang memiliki 19 blok dan *Cluster* rendah memiliki 12 blok [29]. Kedua penelitian terdahulu tersebut telah berhasil menerapkan Algoritma *K-Means* dalam mengelompokkan dan memperoleh hasil sebagai tolak ukur keberhasilan dalam melakukan penelitian, dengan adanya penelitian terdahulu tersebut maka penelitian ini juga akan menerapkan Algoritma *K-Means* untuk menyelesaikan masalah pengelompokan hasil produksi tanaman perkebunan berdasarkan provinsi di Indonesia.

Berdasarkan latar belakang yang telah diuraikan, maka perlu dilakukan penelitian ini, diharapkan dengan dilakukannya penelitian ini maka provinsi yang masuk dalam *Cluster* terendah segera mendapat perhatian pemerintah agar hasil produksi tanaman perkebunan mereka jangan sampai menurun untuk tahun-tahun yang akan datang.

## 2. METODE PENELITIAN

### 2.1. Data Penelitian

Dataset penelitian yang digunakan pada penelitian ini adalah data hasil produksi tanaman perkebunan berdasarkan 34 Provinsi yang ada di Indonesia, yang diperoleh secara langsung dari situs resmi Badan Pusat Statistik. Dataset penelitian dapat dilihat pada tabel 1 berikut.

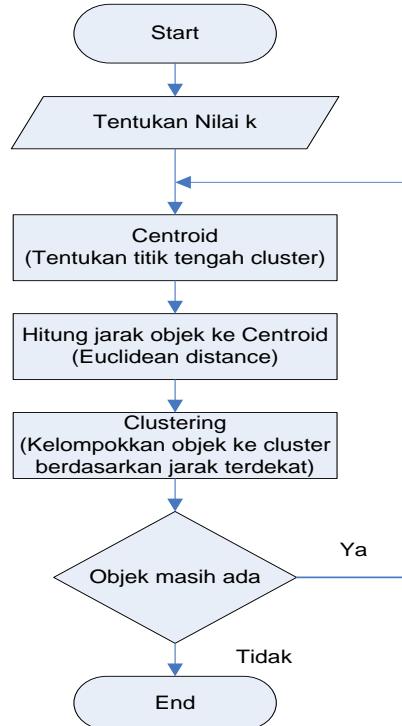
Tabel 1. Produksi Tanaman Perkebunan (Ribu Ton)

Provinsi	Kelapa Sawit			Kelapa			Karet		
	2018	2019	2020	2018	2019	2020	2018	2019	2020
Aceh	1.037,40	1.133,30	1.134,60	63,50	63,80	63,60	93,70	85,20	74,80
Sumatera Utara	5.737,30	5.647,30	5.776,80	99,40	100,50	100,80	418,90	387,70	327,70
Sumatera Barat	1.248,30	1.253,40	1.312,30	78,00	78,30	77,60	152,50	142,00	132,10
Riau	8.496,00	9.512,90	9.984,30	392,70	390,70	399,40	337,30	308,00	291,90
Jambi	2.691,30	2.884,40	3.022,60	107,90	108,90	109,60	319,50	301,40	262,80
Sumatera Selatan	3.793,60	4.049,20	4.267,00	57,70	57,60	55,40	1.043,00	944,20	804,80
Bengkulu	1.047,70	1.032,10	1.063,40	9,20	8,80	9,50	126,30	113,60	94,10
Lampung	487,20	414,20	384,90	86,90	83,40	83,40	174,10	148,50	136,90
Kep. Bangka Belitung	900,30	815,70	843,00	4,40	4,80	5,10	59,90	55,10	46,50
Kep. Riau	28,90	22,80	20,00	11,60	11,90	12,50	29,40	23,30	19,00
DKI Jakarta	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Jawa Barat	46,00	32,20	33,10	93,60	87,90	87,60	67,50	44,90	40,70
Jawa Tengah	0,00	0,00	0,00	172,60	169,00	167,90	38,50	29,50	30,90
DI Yogyakarta	0,00	0,00	0,00	47,40	48,10	46,50	0,00	0,00	0,00
Jawa Timur	0,00	0,00	0,00	244,10	240,40	240,10	27,40	22,60	24,00

Provinsi	Kelapa Sawit			Kelapa			Karet		
	2018	2019	2020	2018	2019	2020	2018	2019	2020
Banten	38,40	31,30	27,40	43,30	43,10	43,60	17,60	15,80	12,70
Bali	0,00	0,00	0,00	66,40	66,90	67,30	0,40	0,10	0,00
Nusa Tenggara Barat	0,00	0,00	0,00	45,60	47,10	47,40	0,00	0,00	0,00
Nusa Tenggara Timur	0,00	0,00	0,00	69,60	70,10	70,40	0,00	0,00	0,00
Kalimantan Barat	3.086,90	5.235,30	5.471,40	83,70	83,80	84,80	272,30	261,50	236,00
Kalimantan Tengah	7.230,10	7.664,80	7.685,80	15,70	15,60	15,70	161,90	152,20	125,90
Kalimantan Selatan	1.464,20	1.665,40	1.561,10	25,20	24,90	24,20	188,40	174,60	145,10
Kalimantan Timur	3.786,50	3.988,90	3.823,20	10,70	9,50	12,50	79,90	76,90	65,50
Kalimantan Utara	305,10	281,40	301,60	0,50	0,50	0,50	0,80	0,80	0,60
Sulawesi Utara	0,00	0,00	0,00	262,50	271,80	250,70	0,00	0,00	0,00
Sulawesi Tengah	383,60	381,70	371,70	193,90	195,70	195,70	4,60	3,70	3,60
Sulawesi Selatan	105,70	91,00	100,30	72,10	70,90	54,80	10,90	5,30	4,60
Sulawesi Tenggara	106,10	59,50	76,30	42,80	41,10	41,30	0,10	0,00	0,00
Gorontalo	9,90	16,20	5,00	58,80	60,90	60,60	0,00	0,00	0,00
Sulawesi Barat	386,20	348,40	348,00	36,60	37,30	37,20	0,00	0,00	0,00
Maluku	23,60	17,60	19,10	103,00	104,30	103,80	1,40	0,60	0,70
Maluku Utara	0,00	0,00	0,00	209,80	210,90	211,40	0,00	0,00	0,00
Papua Barat	98,10	103,50	106,40	15,80	16,20	16,20	0,00	0,00	0,00
Papua	345,10	437,70	557,60	15,20	15,20	14,80	4,10	4,10	3,70

## 2.2. Flowchart Penelitian

Flowchart penelitian dengan algoritma K-Means pada penelitian ini dapat dilihat pada gambar 1 berikut.



Gambar 1. Flowchart Penelitian [30]

Tahapan demi tahapan metode *K-Means* sebagai berikut [31]–[35]:

1. Atur jumlah *Cluster* (*k*) pada data set.
2. Atur nilai pusat (*Centroid*).

Pada tahap awal Pengaturan nilai *Centroid* dilakukan secara acak. Pada tahap iterasi digunakan rumus persamaan (1) seperti berikut.

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (1)$$

Keterangan :

- $V_{ij}$  = *Centroid* rata-rata *Cluster* ke-*I* variabel ke-*j*
- $N_i$  = Jumlah anggota *Cluster* ke-*i*
- i, k* = Indeks *Cluster*
- j* = Indeks variabel
- $X_{kj}$  = nilai ke-*k* variabel ke-*j* untuk *Cluster*

3. Berdasarkan masing-masing *record*, jarak terdekat dihitung dengan *Centroid*.

Beberapa cara yang biasa digunakan untuk melakukan pengukuran jarak data ke pusat *Cluster*, diantaranya *Euclidean*, *Manhattan/City Block*, dan *Minkowsky*. Setiap cara memiliki kelebihan dan kekurangan masing-masing. Untuk penulisan pada bab ini, jarak *Centroid* yang digunakan adalah *Euclidean Distance*, dengan rumus seperti berikut ini:

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (2)$$

Keterangan :

*De* = *Distance Euclidean*

*i* = Jumlah objek<sup>2</sup>

(*x, y*) = Koordinat objek

(*s, t*) = Koordinat *Centroid*

4. Kelompokkan objek berdasarkan jarak ke *Centroid* terdekat

5. Ulangi langkah ke-3 hingga langkah ke-4, lakukan *iterasi* hingga *Centroid* bernilai optimal.

### 3. HASIL DAN ANALISIS

#### 3.1. Perhitungan Manual Algoritma K-Means *Clustering*

1. Menentukan Data *Cluster*

Data yang akan di *Cluster* adalah data hasil produksi tanaman perkebunan di Indonesia berdasarkan provinsi (tabel 1).

2. Menentukan Nilai k Jumlah *Cluster*

Jumlah *Cluster* sebanyak 3 *Cluster*, *Cluster* yang akan dibentuk diantaranya yaitu *Cluster* rendah (C1), *Cluster* sedang (C2) dan *Clustering* tinggi (C3).

3. Menentukan Nilai *Centroid* (Pusat *Cluster*)

Penentuan pusat *Cluster* awal ditentukan secara random yang diambil dari data yang ada dalam range. Adapun nilai untuk *Cluster* rendah (C1) diambil dari nilai rendah, *Cluster* sedang (C2) diambil dari nilai pertengahan dan *Cluster* tinggi (C3) diambil dari nilai yang paling tinggi. Daftar *Centroid* data awal (pusat *Cluster*) dapat dilihat pada tabel 2 berikut.

Tabel 2. Pusat *Cluster* Awal

Cluster 3	9331.067	394.2667	312.4
Cluster 2	4597.867	84.1	256.6
Cluster 1	0	70.03333	0.133333

4. Menghitung Jarak dari Setiap *Centroid*

Setelah data nilai pusat *Cluster* awal ditentukan, maka langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat *Cluster*. Proses pencarian data terpendek pada iterasi ke-1 dapat dilihat pada perhitungan berikut:

$$C_{1,3} = \sqrt{1101.767 - 9331.067)^2 + (63.6333 - 394.2667)^2 + (84.5666 - 312.4)^2} = 8239.09$$

$$C_{1,2} = \sqrt{1101.767 - 4597.867)^2 + (63.6333 - 84.1)^2 + (84.5666 - 256.6)^2} = 3500.90$$

$$C_{1,1} = \sqrt{(1101.767 - 0)^2 + (63.6333 - 70.03333)^2 + (84.5666 - 0.133333)^2} = 1105.016$$

Begitu seterusnya hingga C34,3

Jarak hasil perhitungan dilakukan perbandingan dan dipilih jarak terdekat antara titik tiap objek dengan titik *Centroid*, jarak menunjukkan bahwa data berada dalam satu kelompok dengan *Centroid* terdekat. Tabel jarak terdekat dari *Centroid* iterasi ke-1 seperti pada tabel 3 berikut.

Tabel 3. Hasil Perhitungan Algoritma K-Means Iterasi ke-1

C1	C2	C3	Jarak
1105.016	3500.39	8239.09	1105.016
3846.733	779.5625	5511.308	779.5625
1279.271	3328.506	8067.734	1279.271
9341.91892	4743.68	0	0
2881.444616	1732.359	6471.291	1732.359
4142.487706	877.5685	290788.7	877.5685
1055.374162	3553.895	8294.718	1055.374
455.4901892	4170.383	8909.109	455.4902
857.1770101	3751.191	8490.947	857.177
67.11133921	4580.45	9319.481	67.11134
308.4716495	4305.07	9044.234	308.4716
65.98482859	4565.401	9302.628	65.98483
105.0621817	4604.101	9337.948	105.0622
27.39377277	4589.859	9327.425	27.39377
173.2458494	4606.404	9336.752	173.2458
44.63794219	4572.051	9310.06	44.63794
3.166835441	4605.045	9342.028	3.166835

C1	C2	C3	Jarak
5866666.867	5866412	5866362	5866362
0	4605.036	9341.919	0
4605.035719	0	4743.68	0
7528.522522	2931.894	1850.898	1850.898
1573.350173	3036.133	7777.599	1573.35
3867.359598	757.6286	5483.478	757.6286
304.0903791	4310.247	9048.979	304.0904
1276.906862	3348.329	8075.048	1276.907
399.1207416	4227.882	8959.593	399.1207
99.31792386	4505.826	9242.953	99.31792
85.45544984	4524.712	9262.417	85.45545
14.77798673	4594.532	9331.798	14.77799
362.3724399	4245.022	8982.744	362.3724
39.21788428	4584.945	9320.706	39.21788
190.6813896	4478.284	9209.464	190.6814
115.9869078	4503.006	9241.413	115.9869
450.1847045	4159.32	8897.703	450.1847

## 5. Menentukan Posisi *Cluster*

Dalam menentukan *Cluster* dengan mencari nilai *Cluster* berdasarkan nilai *Cluster* dan diletakkan pada *Cluster* yang sesuai dengan nilai minimal pada iterasi ke-1. *Cluster* iterasi ke-1 dapat dilihat seperti pada tabel 4 berikut.

Tabel 4. Hasil Perhitungan Jarak Pusat *Cluster* Iterasi 1

Proses K-Means akan terus ber iterasi sampai pengelompokan data sama dengan pengelompokan data iterasi sebelumnya. Dengan kata lain, proses akan terus melakukan iterasi sampai data pada iterasi terakhir sama dengan iterasi sebelumnya.

Langkah selanjutnya adalah mencari *Cluster* pada iterasi selanjutnya sampai nilai iterasi nya sama. Berikut adalah proses perhitungan iterasi ke-2.

1. Menentukan nilai *Centroid* baru.

Dalam mendapatkan nilai *Centroid* baru untuk melanjutkan iterasi selanjutnya dengan mencari nilai rata-rata dari masing-masing *Cluster*. Berikut ini adalah perhitungan *Centroid* baru pada masing-masing *Cluster* :

$$C_{1,a} = (1101.8 + 1271.3 + 1047.7 + 428.77 + 853 + 23.9 + 300.1 + 0 + 15.333 + 0 + 32.367 + 0 + 12.8 + 0 + 1563.6 + 296.03 + 1262.2 + 379 + 99 + 80.633 + 10.367 + 360.87 + 20.1 + 128.73 + 102.67 + 446.8) / 24 = 411.42$$

$$\begin{aligned}
 C_{1,b} &= (63.63 + 77.967 + 9.1667 + 84.567 + 4.7667 + 12 + 1.4667 + 89.7 + 169.83 + 47.333 + 241.53 + 43.333 + 66.867 + 46.7 + 70.033 + 24.767 + 0.5 + 261.67 + 195.1 + 65.933 + 41.733 + 60.1 + 37.033 + 103.7 + 210.7 + 16.067 + 15.067) / 24 = 85.88 \\
 C_{1,c} &= (84.567 + 142.2 + 111.33 + 153.17 + 53.833 + 23.9 + 19.967 + 51.033 + 32.967 + 0 + 24.667 + 15.367 + 0.1667 + 58.667 + 0.1333 + 169.37 + 0.7333 + 26.633 + 3.9667 + 6.9333 + 0.0333 + 3.6333 + 0.0333 + 0.9 + 0 + 0.4667 + 3.9667) / 24 = 2483.2 \\
 C_{2,a} &= (3828 + 2866.1 + 4036.6 + 4597.9 + 3866.2) / 5 = 3839 \\
 C_{2,b} &= (100.23 + 108.8 + 56.9 + 84.1 + 10.9) / 5 = 72.187 \\
 C_{2,c} &= (378.1 + 294.57 + 930.67 + 256.6 + 74.1) / 5 = 386.81 \\
 C_{3,a} &= (9331.1 + 7526.9) / 2 = 8429 \\
 C_{3,b} &= (394.27 + 15.667) / 2 = 204.97 \\
 C_{3,c} &= (312.4 + 146.67) / 2 = 229.53
 \end{aligned}$$

Maka nilai *Centroid* baru pada iterasi ke-2 dapat dilihat seperti pada tabel 5 berikut.

Tabel 5. *Centroid* Baru Iterasi Ke-2

<i>Cluster 3</i>	8429	204.97	229.53
<i>Cluster 2</i>	2839	72.187	386.81
<i>Cluster 1</i>	411.42	85.88	2483.2

## 2. Menghitung jarak dari *Centroid*

Setelah mendapatkan nilai *Centroid* baru, maka dilakukan perhitungan untuk iterasi ke-2. Proses perhitungan jarak dari *Centroid* baru untuk iterasi ke-2 adalah sebagai berikut:

$$C_{1,3} = \sqrt{(1101.767 - 8429)^2 + (63.6333 - 204.97)^2 + (84.5666 - 229.53)^2} = 8239.09$$

$$C_{1,2} = \sqrt{(1101.767 - 3839)^2 + (63.6333 - 72.187)^2 + (84.5666 - 389.81)^2} = 2753.882$$

$$C_{1,1} = \sqrt{(1101.767 - 411.42)^2 + (63.6333 - 85.88)^2 + (84.5666 - 2483.2)^2} = 2496.1$$

Begitu seterusnya sampai dengan  $C_{30,2}$ .

Tabel jarak terdekat antara titik tiap objek dengan titik *Centroid* dapat dilihat seperti pada tabel 6 berikut.

Tabel 6. Jarak Terdekat Iterasi Ke-2

C1	C2	C3	Jarak
2496.100175	2753.882	8239.09	2496.1
4013.062537	31.36015	4604.589	31.36015
2493.951481	2579.299	7159.326	2493.951
9185.182502	5502.006	925.4326	925.4326
3288.783202	977.9487	5564.111	977.9487
3943.745621	578.8433	213661.1	578.8433
2537.286401	845.4992	8321.104	845.4992
2357.986741	234.0208	8841.165	234.0208
2537.024707	652.3154	8498.529	652.3154
2496.181629	200.6249	9243.474	200.6249
2494.19086	135.6306	8997.717	135.6306
2457.79136	172.3407	9198.43	172.3407
2470.507292	233.1251	9200.764	233.1251
2515.433733	209.6548	9237.737	209.6548
2473.494588	272.8332	9172.752	272.8332
2499.41012	188.6705	9223.199	188.6705
2514.060916	222.2336	9243.709	222.2336
5864183.814	5866581	5866263	5864184
2513.692812	222.1932	9242.419	222.1932
4912.407663	4396.229	5366.64	4396.229
7671.293512	7322.4	3937.227	3937.227
2695.283776	1361.986	7857.553	1361.986
4378.406083	3661.762	5965.507	3661.762
2513.224793	143.8107	9002.66	143.8107
2667.718093	1075.676	8019.281	1075.676
2491.122623	228.2627	8845.707	228.2627
2500.379276	132.2926	9153.608	132.2926
2511.446759	154.132	9180.615	154.132
2510.582173	211.6154	9236.87	211.6154
2511.11723	181.409	8928.548	181.409
2507.154859	205.8922	9210.352	205.8922
2491.476861	179.9257	9067.98	179.9257
2513.465678	144.606	9171.234	144.606
2516.346017	261.6377	8860.248	261.6377

## 3. Menentukan *Cluster* Pengelompokan

Penentuan anggota tiap-tiap *Cluster* dengan menghitung jarak objek yang memiliki nilai minimum dan diletakkan pada *Cluster* yang sesuai dengan nilai minimum. Anggota tiap-tiap *Cluster* iterasi ke-2 dapat dilihat pada tabel 7.

Tabel 7. *Cluster* Iterasi Ke-2

Setelah melihat hasil dari iterasi ke-2 kemudian dibandingkan dengan iterasi ke -1 memiliki hasil yang berbeda. Maka perhitungan manualnya dilanjutkan pada iterasi ke-3.

### 1. Menentukan nilai *Centroid* baru

Dalam mendapatkan nilai *Centroid* baru untuk melanjutkan iterasi selanjutnya dengan mencari nilai rata-rata dari masing-masing *Cluster* berdasarkan dataset pada tabel 4.7. berdasarkan *Cluster* yang terpilih dari nilai tertinggi sebagai berikut :

$$C_{1,a} = (2496.10018 + 249395148) / 2 = 2495.026$$

$$C_{1,b} = (2753.88 + 2579.3) / 2 = 2666.591$$

$$C_{1,c} = (8239.09 + 7159.33) / 2 = 7699.208$$

$$C_{2,a} = (4013.06254 + 3288.7832 + 3943.74562 + 2537.2864 + 2357.98674 + 2537.02471 + 2469.18163 + 2494.19086 + 2457.79136 + 2470.50729 + 2515.43373 + 2473.49459 + 2499.41012 + 2514.06092 + 5864183.81 + 2513.69281 + 4912.40766 + 2695.28378 + 4378.40608 + 2513.22479 + 2667.71809 + 2491.12262 + 2500.37928 + 2511.44676 + 2510.58217 + 2511.11723 + 507.15486 + 2491.47686 + 2513.46568 + 2516.34602) / 30 = 198167.2$$

$$C_{2,b} = \frac{(31.3602 + 977.949 + 578.843 + 845.499 + 234.021 + 652.315 + 200.625 + 135.631 + 172.341 + 233.125 + 209.655 + 272.833 + 188.671 + 222.234 + 5866581 + 222.193 + 4396.23 + 1361.99 + 3661.76 + 143.811 + 1075.68 + 228.263 + 132.293 + 154.132 + 211.615 + 181.409 + 205.892 + 179.926 + 144.606 + 261.638)}{30} = 196136.6$$

$$C_{2,C} = \frac{(8239.09 + 7159.33 + 5564.11 + 213661 + 8321.1 + 8841.16 + 8498.53 + 9243.47 + 8997.72 + 9198.43 + 9200.76 + 9237.74 + 9172.75 + 9223.2 + 9243.71 + 5866263 + 9242.42 + 5366.64 + 7857.55 + 5965.51 + 9002.66 + 8019.28 + 8845.71 + 9143.61 + 9180.62 + 9236.87 + 8928.55 + 9210.35 + 9067.98 + 9171.23 + 8860.25)}{30} = 210546$$

$$C_{3,a} = (9185.1825 + 7671.29351) / 2 = 8428.238$$

$$C_{3,b} = (5502.01 + 7322.4) / 2 = 6412.203$$

$$C_{3,C} = (925.433 + 3937.23) / 2 = 2431.33$$

## 2. Menghitung jarak dari *Centroid*

Setelah mendapatkan nilai *Centroid* baru, maka dilakukan perhitungan untuk iterasi ke-3. Proses perhitungan jarak dari *Centroid* ke-2 untuk iterasi ke-3 adalah sebagai berikut:

$$C_{1,3} = \sqrt{(1101.767 - 8428.238)^2 + (63.6333 - 6412.203)^2 + (84.5666 - 2431.33)^2} = 9974.41$$

$$C_{1,2} = \sqrt{(1101.767 - 198167.2)^2 + (63.6333 - 196136.6)^2 + (84.5666 - 210546)^2} = 348674$$

$$C_{1,1} = \sqrt{(1101.767 - 2495.026)^2 + (63.6333 - 2666.591)^2 + (84.5666 - 7699.208)^2} = 8166,97$$

Tabel jarak terdekat antara titik tiap objek dengan titik *Centroid* dapat dilihat seperti pada tabel 8 berikut.

Tabel 8. Jarak Terdekat Iterasi Ke-3

C1	C2	C3	Jarak
8166.965	348673.8	9974.408	8166.965
7871.571	346941.7	8075.822	7871.571
8081.26	348535.1	9827.714	8081.26
10317.94	343762	6443.642	6443.642
7842.748	347527.1	8673.87	7842.748
7416.204	346514.8	2262771	7416.204
210678.1	196217.4	198309.8	196217.4
210606.2	196159.4	198257.9	196159.4
210685.9	196228	198322.2	196228
210687.6	196246.4	198348.7	196246.4
210695.3	196248.2	198347.8	196248.2
210608.8	196168	198270.2	196168
210529.8	196089.3	198192	196089.3
210653.3	196211.7	198314.1	196211.7
210458.5	196017.8	198120.4	196017.8
210656.5	196214.9	198317.2	196214.9
210633.9	196192.7	198295.2	196192.7
5862748	5867282	5867524	5862748
210630.8	196189.5	198292.1	196189.5
210604.6	196075.7	198132.8	196075.7
210727.4	196140.3	198167.5	196140.3
210657.6	196187.7	198275	196187.7
210678	196159.4	198223.6	196159.4
210697	196249.5	198349.2	196249.5
210427.5	195960.4	198050.6	195960.4
210501.6	196052.4	198151.3	196052.4
210633.5	196190.3	198291.9	196190.3
210658.1	196215.2	198316.9	196215.2
210640.4	196199.1	198301.5	196199.1
210659.9	196211	198310	196211
210596.9	196155.2	198257.6	196155.2
210488.8	196044.8	198146.1	196044.8
210683.5	196240.1	198341.6	196240.1
210680.9	196230.3	198328.5	196230.3

### 3. Menentukan *Cluster* pengelompokan

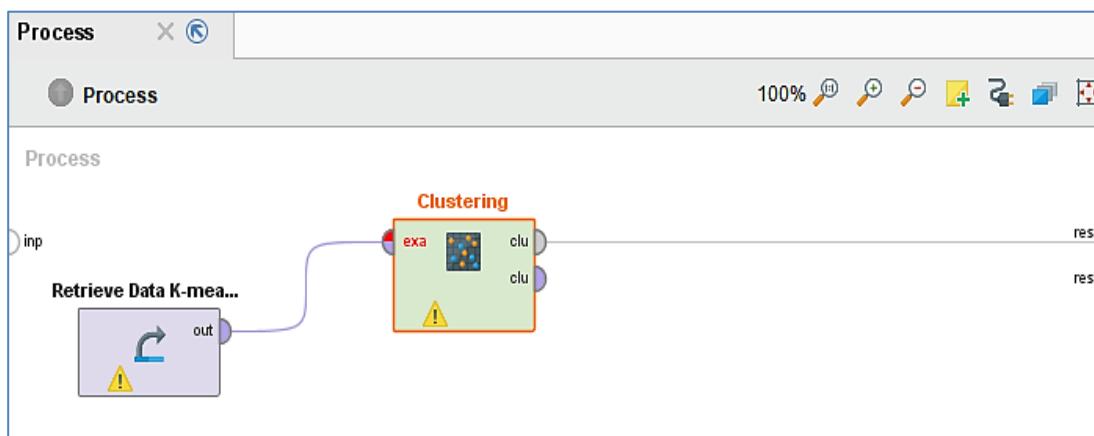
Dalam menentukan anggota tiap-tiap *Cluster* dengan menghitung jarak objek yang memiliki nilai minimum dan diletakkan pada *Cluster* yang sesuai dengan nilai minimum. Anggota tiap-tiap *Cluster* iterasi ke-3 dapat dilihat seperti pada tabel 9 berikut.

Tabel 9. Cluster Iterasi Ke-3

Berdasarkan tabel *Cluster* iterasi ke-5 dan tabel *Cluster* iterasi ke-6 memiliki nilai *Cluster* yang sama dan tidak ada lagi perpindahan dari *Cluster* satu ke *Cluster* yang lain. Maka perhitungannya dihentikan pada iterasi ke- 6 dan hasil yang diperoleh dari ke enam iterasi tersebut yaitu C1 sebanyak 27 Data dan C2 sebanyak 2 Data dan C3 sebanyak 6 Data.

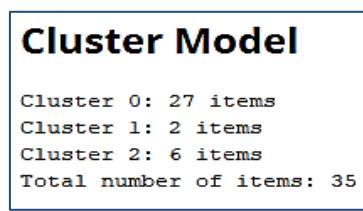
### **3.2. Pengujian dengan Rapid Miner**

Berikut adalah proses pengelompokan dan hasil dari metode K-Means yang dilakukan dengan Rapid miner. Tahapan pertama dengan meng klik *Clustering* and *Segmentation* lalu pilih K-Means dapat dilihat pada gambar 2. Lalu hubungkan antara read excel dengan *Clustering* seperti gambar 2 berikut.



Gambar 2. Proses K-Means dengan Rapid Miner

Untuk mendapatkan hasil pengelompokan maka pada tahap selanjutnya klik run pada toolbar. Pada tahap ini akan menampilkan hasil akhir serta langkah terakhir dalam penggunaan Tolls Rapid Miner seperti yang terlihat pada gambar 3 berikut.



Gambar 3. Hasil *Cluster*

Berdasarkan gambar 3 dapat dijelaskan bahwa dari 3 (tiga) *Cluster* yang dihasilkan terdapat 27 items untuk *Cluster* 0 (*Cluster* tinggi), 2 items untuk *Cluster* 1 (*Cluster* Sedang), dan 6 items untuk *Cluster* 2 (*Cluster* rendah). Untuk hasil akhir tabel *Centroid* dapat dilihat pada gambar 4. Sedangkan hasil pengelompokan hasil produksi tanaman perkebunan berdasarkan provinsi menggunakan Rapid Miner dapat dilihat pada gambar 4. Berdasarkan gambar 4 dapat diketahui bahwa pada kelompok tinggi memiliki 6 node yaitu warna merah, pada kelompok sedang memiliki 2 node berwarna jingga, sedangkan pada kelompok rendah memiliki 27 node dengan warna biru.



Gambar 4. Plot View Hasil Cluster

#### 4. KESIMPULAN

Penerapan Data Mining menggunakan Algoritma K-Means dapat mengelompokkan data produksi tanaman perkebunan di Indonesia. Penelitian ini menghasilkan 3 *Cluster* yaitu *Cluster* rendah, *Cluster* sedang, dan *Cluster* tinggi. Berdasarkan analisis perhitungan menggunakan Algoritma K-Means diperoleh hasil 3 *Cluster* dengan *Cluster* tinggi sebanyak 6 Provinsi. *Cluster* sedang sebanyak 2 Provinsi dan *Cluster* rendah sebanyak 27 Provinsi. Berdasarkan hasil pengelompokan menggunakan Algoritma K-Means pada data produksi tanaman perkebunan di Indonesia menggunakan aplikasi Rapid Miner 8, diperoleh perhitungan yang sama dengan perhitungan manual. Dengan demikian hasil penelitian ini dapat menjadi masukan kepada pihak pemerintah Indonesia, khususnya pemerintah provinsi masing-masing dalam upaya peningkatan produksi tanaman perkebunan untuk tahun-tahun yang akan datang.

#### REFERENSI

- [1] M. A. Amri, D. Hartama, A. Wanto, Sumarno, and H. S. Tambunan, "Penerapan Metode Fuzzy Mamdani dalam Penentuan Penerima BLT-DD di Mekar Sari Raya," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 4, no. 1, pp. 269–277, 2020.
- [2] A. Wulandari, R. P. Saragih, Maslina Manurung, A. Wanto, and rfan S. Damanik, "Sistem Pendukung Keputusan pada Pemilihan Masker Wajah Facial mask Berdasarkan Konsumen dengan Metode Analytical Hierarchy Process," in *Seminar Nasional Ilmu Sosial dan Teknologi (SANISTEK)*, 2021, pp. 197–201.
- [3] F. Fania, M. Azzahra, D. Hartama, A. Wanto, and A. Rahim, "Rekomendasi Pemilihan Calon Peserta MTQ Terbaik Tahun 2019 dengan Teknik Additive Ratio Assessment (ARAS)," in *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, 2021, pp. 608–612.
- [4] D. N. Batubara, A. Padillah, Chairunnisa, A. Wanto, and Saifullah, "Penerapan Metode VIKOR Untuk Menentukan Susu Lansia Terbaik," in *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, 2021, pp. 586–591.
- [5] S. R. Ningsih, D. Hartama, A. Wanto, I. Parlina, and Solikhun, "Penerapan Sistem Pendukung Keputusan Pada Pemilihan Objek Wisata di Simalungun," in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2019, pp. 731–735.
- [6] M. Widyasuti, A. Wanto, D. Hartama, and E. Purwanto, "Rekomendasi Penjualan Aksesoris Handphone Menggunakan Metode Analytical Hierarchy Process (AHP)," *Konferensi Nasional Teknologi Informasi dan Komputer*, vol. 1, no. 1, pp. 27–32, 2017.
- [7] A. Wanto and H. Damanik, "Analisis Penerapan Sistem Pendukung Keputusan Terhadap Seleksi Penerima Beasiswa BBM (Bantuan Belajar Mahasiswa) Pada Perguruan Tinggi Menggunakan Metode Simple Additive Weighting (SAW) (Studi Kasus : AMIK Tunas Bangsa Pematangsiantar)," in *Seminar Nasional Rekayasa (SNTR) II*, 2015, no. 2, pp. 323–333.
- [8] A. Wanto and E. Kurniawan, "Seleksi Penerimaan Asisten Laboratorium Menggunakan Algoritma AHP Pada AMIK-STIKOM Tunas Bangsa Pematangsiantar," *Jurnal Informatika dan Komputer (JIKO)*, vol. 3, no. 1, pp. 11–18, 2018.

- [9] S. P. Siregar, D. Hartama, and A. Wanto, "Estimasi Nilai Tukar Petani Subsektor Tanaman Pangan Menggunakan JST pada Provinsi Sumatera Utara," *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, vol. 1, no. September, pp. 369–377, Sep. 2019.
- [10] C. Astria, H. Okprana, A. Wanto, D. Hartama, and H. S. Tambunan, "Implementasi Inferensi Fuzzy Tsukamoto pada Prediksi Penjualan Telur Ayam Eropa pada Bisnis Raffa Telur," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 4, no. 1, pp. 58–61, 2020.
- [11] S. Andini, R. Sitanggang, A. Wanto, H. Okprana, G. S. Achmad Daengs, and S. Aryza, "Utilization of Rough Sets Method with Optimization Genetic Algorithms in Heart Failure Cases," *Journal of Physics: Conference Series*, vol. 1933, no. 1, pp. 1–9, 2021.
- [12] W. Saputra, J. T. Hardinata, and A. Wanto, "Penerapan Metode Resilient dalam Menentukan Model Arsitektur Terbaik untuk Prediksi Pengangguran Terbuka di Indonesia," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, 2019, pp. 21–29.
- [13] R. D. Dana, A. R. Dikananda, D. Sudrajat, A. Wanto, and F. Fasya, "Measurement of health service performance through machine learning using Clustering techniques," *Journal of Physics: Conference Series*, vol. 1360, no. 1, pp. 1–6, 2019.
- [14] Y. Andriani, A. Wanto, and H. Handrizal, "Jaringan Saraf Tiruan dalam Memprediksi Produksi Kelapa Sawit di PT. KRE Menggunakan Algoritma Levenberg Marquardt," *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, vol. 1, no. September, pp. 249–259, 2019.
- [15] I. C. Saragih, D. Hartama, and A. Wanto, "Prediksi Perkembangan Jumlah Pelanggan Listrik Menurut Pelanggan Area Menggunakan Algoritma Backpropagation," *Building of Informatics, Technology and Science (BITS)*, vol. 2, no. 1, pp. 48–54, 2020.
- [16] M. Syafiq, D. Hartama, I. O. Kirana, I. Gunawan, and A. Wanto, "Prediksi Jumlah Penjualan Produk di PT Ramayana Pematangsiantar Menggunakan Metode JST Backpropagation," *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, p. 175, 2020.
- [17] M. Situmorang, A. Wanto, and Z. M. Nasution, "Architectural Model of Backpropagation ANN for Prediction of Population-Based on Sub-Districts in Pematangsiantar City," *International Journal of Information System & Technology*, vol. 3, no. 1, pp. 98–106, 2019.
- [18] N. Z. Purba, A. Wanto, and I. O. Kirana, "Implementation of ANN for Prediction of Unemployment Rate Based on Urban Village in 3 Sub-Districts of Pematangsiantar," *International Journal of Information System & Technology (IJISTECH)*, vol. 3, no. 1, pp. 107–116, 2019.
- [19] I. I. P. Damanik, S. Solikhun, I. S. Saragih, I. Parlina, D. Suhendro, and A. Wanto, "Algoritma K-Medoids untuk Mengelompokkan Desa yang Memiliki Fasilitas Sekolah di Indonesia," in *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2019, vol. 1, no. September, pp. 520–527.
- [20] F. S. Napitupulu, I. S. Damanik, I. S. Saragih, and A. Wanto, "Algoritma K-Means untuk Pengelompokan Dokumen Akta Kelahiran pada Tiap Kecamatan di Kabupaten Simalungun," *Building of Informatics, Technology and Science (BITS) Volume*, vol. 2, no. 1, pp. 55–63, 2020.
- [21] S. F. Damanik, A. Wanto, and I. Gunawan, "Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok," *Jurnal Krisnadana Volume*, vol. 1, no. 2, pp. 21–32, 2022.
- [22] I. Parlina *et al.*, "Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–5, 2019.
- [23] N. A. Febriyati, A. D. GS, and A. Wanto, "GRDP Growth Rate Clustering in Surabaya City uses the K- Means Algorithm," *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 276–283, 2020.
- [24] M. A. Hanafiah and A. Wanto, "Implementation of Data Mining Algorithms for Grouping Poverty Lines by District/City in North Sumatra," *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 315–322, 2020.
- [25] A. Pradipta, D. Hartama, A. Wanto, S. Saifullah, and J. Jalaluddin, "The Application of Data Mining in Determining Timely Graduation Using the C45 Algorithm," *IJISTECH (International Journal of Information System & Technology)*, vol. 3, no. 1, pp. 31–36, 2019.
- [26] N. Arminarahmah, A. D. GS, G. W. Bhawika, M. P. Dewi, and A. Wanto, "Mapping the Spread of Covid-19 in Asia Using Data Mining X-Means Algorithms," *IOP Conference Series: Materials Science and Engineering*, vol. 1071, no. 1, pp. 1–7, 2021.
- [27] T. H. Sinaga, A. Wanto, I. Gunawan, S. Sumarno, and Z. M. Nasution, "Implementation of Data Mining Using C4.5 Algorithm on Customer Satisfaction in Tirta Lihou PDAM," *Journal of Computer Networks, Architecture, and High-Performance Computing*, vol. 3, no. 1, pp. 9–20, 2021.
- [28] D. Haryadi, "Penerapan Algoritma K-Means Clustering Pada Produksi Perkebunan Kelapa Sawit Menurut Provinsi," *Journal of ICT (Informatics and Communications Technology)*, vol. 3, no. 1, pp. 1–15, 2021.
- [29] A. Lili and S. Widodo, "Pengelompokan Hasil Panen Kelapa Sawit Dalam Produksi Per Blok," vol. 01, no. 01, pp. 45–54, 2022.
- [30] Z. S. Younus *et al.*, "Content-based image retrieval using PSO and k-means Clustering algorithm," *Arabian Journal of Geosciences*, vol. 8, no. 8, pp. 6211–6224, 2015.

- [31] E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi Offset, 2012.
- [32] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. New Jersey: John Wiley & Sons, 2005.
- [33] R. Primartha, *Belajar Machine Learning Teori dan Praktik*. Bandung: Informatika Bandung, 2018.
- [34] T. Khotimah, “Pengelompokan Surat dalam Al Qur'an menggunakan Algortima K-Means,” *Jurnal Simetris*, vol. 5, no. 1, pp. 83–88, 2014.
- [35] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, “Memanfaatkan Algoritma K-Means dalam Menentukan Pegawai yang Layak Mengikuti Asessment Center untuk Clustering Program SDP,” *CESS (Journal of Computer Engineering System and Science)*, vol. 3, no. 1, pp. 87–93, 2018.