

Penerapan Algoritma K-Means untuk Klasterisasi Produksi Tanaman Perkebunan di Indonesia

Application of the K-Means Algorithm for Clustering Plantation Crop Production in Indonesia

Reda Maulidina¹
Suastika Yulia Riska^{2*}

^{1,2}Teknik Informatika, Institut Teknologi dan Bisnis ASIA Malang, Indonesia
¹maulidinareda@gmail.com, ²riska.suastika@asia.ac.id

***Penulis Korespondensi:**
Suastika Yulia Riska
riska.suastika@asia.ac.id

Riwayat Artikel:

Diterima : 13 Oktober 2023
Direview : 27 November 2023
Disetujui : 5 Desember 2023
Terbit : 29 Desember 2023

Abstrak

Hasil produksi tanaman perkebunan merupakan salah satu sektor utama dalam menunjang pendapatan masyarakat. Data hasil perkebunan yang dimiliki oleh Badan Pusat Statistik berupa data mentah, yaitu hasil produksi setiap provinsi pada setiap tahunnya. Hal tersebut cukup menyulitkan pemerintah dalam mengetahui provinsi yang berpotensi dalam menghasilkan tanaman perkebunan. *Clustering* hasil produksi tanaman perkebunan, pemerintah akan lebih mudah dalam mengetahui provinsi yang berpotensi menghasilkan tanaman perkebunan. Pada penelitian ini ada 3 produksi tanaman perkebunan yaitu kelapa, kopi, dan kakao. Data yang digunakan adalah data dari tahun 2017 – 2021 yang terdiri dari 29 provinsi. Dari 3 produksi tanaman perkebunan tersebut maka data tersebut dikelompokkan menggunakan teknik Data Mining Algoritma K-Means. Hasil penelitian ini berupa pengelompokan yang terbagi ke dalam 2 *cluster* yang diperoleh dari perhitungan Sum of Squared Error dengan nilai minimum 279261.63 yaitu produksi sedikit, dan produksi banyak. Berdasarkan hasil perhitungan Algoritma K-Means diperoleh produksi pada kelapa diperoleh *cluster* produksi sedikit sebanyak 24 provinsi, *cluster* banyak sebanyak 5 provinsi, untuk kopi diperoleh *cluster* produksi sedikit sebanyak 23 provinsi, *cluster* banyak sebanyak 6 provinsi, dan kakao diperoleh *cluster* produksi sedikit sebanyak 23 provinsi, *cluster* banyak sebanyak 6 provinsi.

Kata Kunci: *Clustering*; K-Means; Tanaman Perkebunan; Sum of Squared Error

Abstract

Plantation crop production is one of the main sectors in increasing people's income. Data on plantation products held by the Central Statistics Agency (BPS) is in the form of raw data, namely the production results of each province each year. This makes it quite difficult for the government to identify provinces that have the potential to produce crops. By clustering plantation crop production results, it will be easier for the government to identify provinces that have the potential to produce plantation crops. In this research there were 3 plantation crop production, namely coconut, coffee and cocoa. The data used is data for 2017 – 2021 which consists of 29 provinces. From the 3 plantation crop production, the data was collected using the K-Means Algorithm Data Mining technique. The results of this research are groupings which are divided into 2 clusters obtained from the Sum of Squared Error (SSE) calculation with a minimum value of 279261.63, namely low production and large production. Based on the results of the K-Means Algorithm calculations, it

was found that coconut production had a small production cluster of 24 provinces, a large cluster of 5 provinces, for coffee a small production cluster of 23 provinces was obtained, a large production cluster was 6 provinces, and for cocoa a small production cluster was obtained of 23 provinces, a cluster of 6 provinces.

Keywords: Clustering; K-Means; Plantation Crops; Sum of Squared Error

1. Pendahuluan

Indonesia merupakan negara yang didukung oleh subsektor pertanian, termasuk perkebunan. Perkebunan merupakan sektor terpenting dalam perekonomian Indonesia [1]. Pasalnya, hasil panen yang tinggi dapat membantu perekonomian Indonesia, terutama melalui transaksi dalam bentuk ekspor [2]. Transaksi dalam bentuk ekspor merupakan kegiatan perekonomian yang paling penting karena merupakan sumber penerimaan devisa negara dan dapat menciptakan lapangan kerja atau menciptakan lapangan kerja serta dapat meningkatkan kesejahteraan masyarakat pada masa pemerintahan dari produksi [2][3]. Setiap provinsi memiliki keunggulan hasil produksi tanaman yang disesuaikan dengan keadaan tanah dan lahan yang dimiliki provinsi tersebut.

Produk tanaman yang sangat penting untuk diekspor di Indonesia adalah kelapa, kopi, dan kakao. Setiap provinsi mempunyai hasil hutan tanaman yang luar biasa, disesuaikan dengan tanah dan kondisi tanah di provinsi tersebut. Berdasarkan data Badan Pusat Statistik (BPS) mengenai hasil produksi tanaman, belum ada data agregat berdasarkan hasil produksi tanaman untuk ketiga komoditas tersebut. Menurut dokumen keterangan yang dihasilkan Direktorat Jenderal Bea dan Cukai melalui situs <https://www.bps.go.id> hasil produksi tanaman perkebunan di Indonesia dari tahun 2017 sampai tahun 2021 pada tanaman perkebunan kopi pada provinsi Aceh dan Sumatera Utara mengalami kenaikan produksi. Hal yang menyebabkan kenaikan hasil produksi disebabkan oleh pengelolaan lahan yang subur, tenaga kerja yang produktif serta luas lahan yang memadai. Hasil produksi tanaman perkebunan kopi yang mengalami kenaikan ditunjukkan pada Tabel 1.

Tabel 1. Kenaikan Hasil Produksi Perkebunan Kopi

Provinsi	Kopi (Ribuan Ton) per Tahun				
	2017	2018	2019	2020	2021
Jawa Tengah	15.70	23.70	24.70	24.90	27.50
Sumatera Utara	67.40	71.00	74.90	75.00	76.80

Sedangkan pada provinsi Kalimantan Selatan mengalami penurunan produksi hasil pada tanaman perkebunan kopi. Hal yang menyebabkan penurunan hasil produksi disebabkan oleh menurunnya luas area perkebunan, kurangnya perawatan dan cara pengelolaan yang kurang tepat serta dukungan pemerintah dalam menunjukkan nilai saing dalam hal ekspor ke Luar Negeri. Hasil produksi tanaman perkebunan kopi yang mengalami penurunan ditunjukkan pada Tabel 2.

Tabel 2. Penurunan Hasil Produksi Perkebunan Kopi

Provinsi	Kopi (Ribuan Ton) per Tahun				
	2017	2018	2019	2020	2021
Kalimantan Selatan	1.60	1.50	1.30	1.30	1.10
Kalimantan Utara	0.20	0.20	0.20	0.20	0.10

Fenomena ini menjadi kewajiban pemerintah untuk membuat kebijakan untuk meningkatkan hasil produksi tanaman perkebunan dalam hal ekspor ke luar negeri. Kurangnya pengolahan data pengelompokan tanaman perkebunan, mengakibatkan terjadinya ketidakseimbangan produksi tanaman perkebunan di Indonesia. Data dari Badan Pusat Statistik (BPS) yang masih

berupa data mentah tersebut harus diolah agar menghasilkan informasi sesuai kebutuhan. Dengan adanya informasi ini pemerintah dapat mengetahui potensi provinsi mana saja yang menghasilkan produksi tanaman perkebunan produksi sedikit dan banyak untuk dijadikan acuan untuk meningkatkan hasil produksi tanaman perkebunan di Indonesia.

Berdasarkan data tersebut, maka perlu dilakukan pengumpulan data secara *cluster* untuk mengetahui provinsi mana saja yang mempunyai potensi produksi perkebunan. *Clustering* diperlukan untuk mengelompokkan data guna mengidentifikasi kelompok (*cluster*) karena pengelompokan elemen-elemen yang lebih kecil berdasarkan kesamaannya satu sama lain [4][5].

Oleh karena itu, metode *clustering* ini berguna untuk menemukan kelompok yang tidak diketahui dalam suatu data [6]. Oleh karena permasalahan diatas maka penelitian ini dilakukan dengan menggunakan teknologi komputer khususnya pada bidang data mining [7][8]. Salah satu jenis pengelompokan dalam data mining adalah *clustering* [6][9][10]. *Clustering* sering dilakukan agar data yang sebelumnya masih dalam keadaan tidak terorganisir dapat dikelompokkan menjadi satu kesatuan sehingga data tersebut dapat dianalisis lebih detail dan menghasilkan wawasan [11].

Salah satu algoritma yang dapat digunakan dalam pengolahan *clustering* data adalah K-Means. Algoritma K-Means digunakan untuk pengelompokan berulang, membagi kumpulan data menjadi beberapa *cluster*[12]. Algoritma K-Means mudah diimplementasikan dan dipahami, proses pembelajarannya relatif cepat dan sangat umum digunakan dalam *clustering* [8]. *Clustering* banyak diterapkan diberbagai bidang kehidupan. *Clustering* dapat digunakan untuk mengelompokkan jumlah wisatawan mancanegara yang ada di Indonesia [13]. Klusterisasi dilakukan dengan menggunakan $k=1$ sampai dengan $k=5$, yang diuji menggunakan Davies Bouldin Index memiliki. Penelitian tersebut, membandingkan metode K-Means dan K-Medoids. Hasilnya menunjukkan bahwa dengan menggunakan Metode K-Means, *cluster* terbaik adalah dengan $k=5$ dengan DBI = 0,302, dan Metode K-Medoids dengan $k=2$ dengan DBI = 0,567 [13]. Penelitian tentang *clustering* menggunakan K-Means dan Metode Elbow juga dapat diterapkan untuk mengukur kesiapan siswa dalam ujian. Pada penelitian tersebut, metode elbow digunakan untuk menentukan *cluster* yang optimal yaitu dikelompokkan menjadi 3 *cluster*. Terdapat 3 *cluster* yang digunakan pada penelitian tersebut adalah siap, cukup siap, dan tidak siap [14]. Metode *clustering* K-Means dan metode optimasi elbow juga dapat digunakan untuk melakukan proses *clustering* pada data pengguna narkoba. Pada penelitian ini menggunakan $k=2$ sampai dengan $k=6$. Hasil perhitungan Metode Elbow menunjukkan bahwa, *cluster* yang optimal adalah pada $k=3$ [15].

Berdasarkan penelitian sebelumnya, menunjukkan bahwa Metode K-Means merupakan salah satu metode *clustering* yang dapat digunakan untuk mempartisi data kedalam dua atau lebih kelompok [16]. Pada penelitian ini dilakukan pengolahan data dengan metode yang dapat mengelompokkan output hasil tanaman perkebunan di Indonesia dalam jumlah besar dan kecil khususnya pada tahun 2017-2021 dengan menggunakan algoritma K-Means. Pada proses optimasi *cluster* yang terbaik menggunakan metode Elbow.

2. Metode Penelitian

Data yang digunakan adalah data statistik produksi tanaman perkebunan yang didapat dari website resmi Badan Pusat Statistik yaitu <https://www.bps.go.id/indicator/54/768/1/produksi-perkebunan-rakyat-menurut-jenis-tanaman.html>. Data tersebut berisikan Provinsi yang ada di Indonesia dan data produksi tanaman perkebunan dari tahun 2017 sampai dengan 2021. Contoh data hasil produksi tanaman perkebunan kelapa, kopi, dan kakao berturut-turut ditunjukkan pada Tabel 3, Tabel 4, dan Tabel 5.

Tabel 3. Data Hasil Produksi Perkebunan Kelapa

Provinsi	Kelapa (Ribu Ton)				
	2017	2018	2019	2020	2021
ACEH	62.80	63.50	63.80	63.60	64.10
SUMATERA UTARA	97.70	99.40	100.50	100.80	100.00
SUMATERA BARAT	69.50	78.00	78.30	77.60	69.10
RIAU	390.60	392.70	390.70	399.40	395.00
...
PAPUA	2.50	2.70	2.80	2.80	2.70

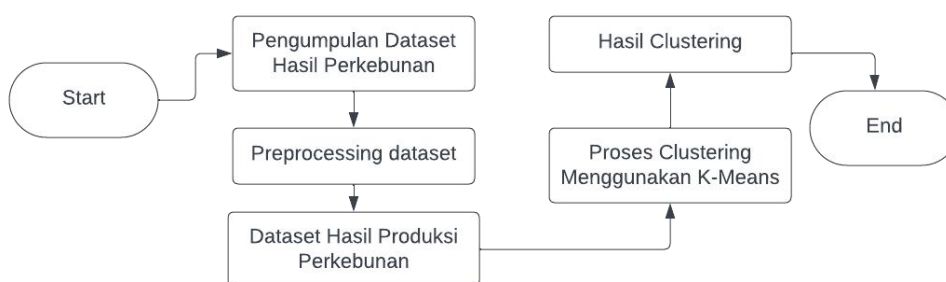
Tabel 4. Data Hasil Produksi Perkebunan Kopi

Provinsi	Kelapa (Ribu Ton)				
	2017	2018	2019	2020	2021
ACEH	68.50	70.80	72.70	73.40	74.20
SUMATERA UTARA	67.40	71.00	74.90	75.00	76.80
SUMATERA BARAT	18.00	18.50	15.30	12.30	12.80
RIAU	2.90	3.00	2.60	2.40	2.40
...
PAPUA	2.50	2.70	2.80	2.80	2.70

Tabel 5. Data Hasil Produksi Perkebunan Kakao

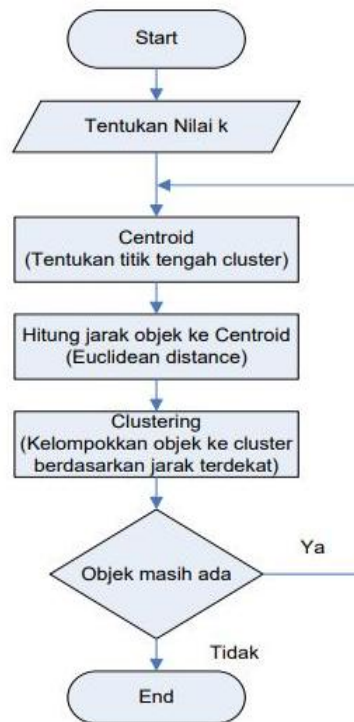
Provinsi	Kelapa (Ribu Ton)				
	2017	2018	2019	2020	2021
ACEH	27.10	39.30	41.10	41.30	40.90
SUMATERA UTARA	24.30	35.40	34.90	35.30	35.90
SUMATERA BARAT	46.20	59.00	53.10	43.30	40.20
RIAU	2.30	3.20	1.10	1.60	1.00
...
PAPUA	9.60	10.80	10.40	10.40	10.50

Diagram pelaksanaan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Flowchart Sistem Clustering Hasil Produksi Perkebunan

Berdasarkan Gambar 1, sebelum proses *clustering* dapat dilakukan, terlebih dahulu dilakukan proses *cleaning* atau pembersihan untuk memastikan data yang akan di-*cluster* sudah benar-benar layak dan sesuai. Proses *preprocessing* dalam penelitian ini mencakup antara lain Menyusun ulang bentuk data dan membuang elemen-elemen yang tidak digunakan atau yang berkemungkinan dapat merusak proses perhitungan. Sehingga, Provinsi yang menunjukkan nilai hasil produksi = 0 dilakukan penghapusan. Selanjutnya dilakukannya proses *clustering* menggunakan K-Means. Proses *Clustering* menggunakan K-Means Ditunjukkan pada Gambar 2.



Gambar 2. Flowchart K-Means

Berdasarkan *flowchart* pada Gambar 2 proses K-Means adalah :

1. Menentukan k sebagai jumlah *Cluster* yang ingin dibentuk
 Hasil akhir dari proses *clustering* adalah data yang sudah dikelompokkan ke dalam sejumlah *cluster*. Dalam *clustering*, penentuan centroid atau pusat awal *cluster* juga mempengaruhi hasil, sehingga pemilihan centroid awal yang tepat dapat meminimalisir kesalahan hasil *cluster*. Untuk mengoptimalkan hasil *clustering*, dalam penelitian ini akan menggunakan metode Elbow yang akan digunakan untuk mencari jumlah *cluster* optimal sekaligus untuk mencari centroid awal terbaik. Ada beberapa tahapan untuk menentukan centroid dan jumlah *cluster* berdasarkan nilai *Sum of Squared Error* (SSE). Tahap pertama adalah pembangkitan beberapa kombinasi centroid dan menghitung nilai *Sum of Squared Error* (SSE) dari centroid tersebut. Tahap kedua adalah menyimpan centroid tahap pertama pada sebuah tabel dan kembali lagi ke tahap pertama sampai jumlah centroid yang diuji terpenuhi. Tahap ketiga adalah memilih satu centroid yang memiliki nilai SSE yang paling minimum dari keseluruhan centroid yang telah dibangkitkan pada tahap pertama. Untuk menghitung nilai SSE digunakan Persamaan 1. sebagai berikut:

$$SSE = \sum_{i=1}^k d(p_i - m_j)^2 \quad (1)$$

Di mana:

- P_i = data ke-i
- M_j = centroid ke-j

1. Perhitungan data ke-1

Perhitungan data ke-1 p1 dengan centroid 1 m1

$$m1 = (62.80 - 53.40)^2 + (63.50 - 47.40)^2 + (63.80 - 48.10)^2 + (63.60 - 46.50)^2 + (64.10 - 46.80)^2$$

$$m1 = 1185.76$$

Perhitungan data ke-1 p1 dengan centroid 2 m2

$$m2 = (62.80 - 253.90)^2 + (63.50 - 244.10)^2$$

$$\begin{aligned}
 &+ (63.80 - 240.40)^2 + (63.60 - 240.10)^2 \\
 &+ (64.10 - 244.50)^2 \\
 m_2 &= 164019.5 \\
 \text{Nilai minimum dari } (p_1, m_1), (p_1, m_2) \\
 \text{min} &= 1185.76
 \end{aligned}$$

2. Menentukan k centroid (titik pusat klaster) awal secara random.
3. Menghitung jarak setiap data ke masing-masing centroid
Selanjutnya proses menghitung jarak antara tiap-tiap data dengan pusat *cluster* (centroid). Untuk menghitung jarak tiap-tiap data terhadap setiap pusat *cluster* (centroid) dapat menggunakan persamaan 2.1 :

$$D(\text{data}, \text{centroid}) = \sqrt{\sum_j^n (\text{data}_i - \text{centroid}_i)^2} \quad (2)$$

Di mana:

- D = euclidean distance
- n = banyaknya objek/jumlah data
- i = atribut ke-i
- j = iterasi

Sebagai contoh, untuk menghitung jarak antara data pertama terhadap tiap-tiap centroid adalah sebagai berikut:

$$D(1,1) = \sqrt{\begin{aligned} &(62.80 - 53.40)^2 + (63.50 - 47.40)^2 + (63.80 - 48.10)^2 \\ &+ (63.60 - 46.50)^2 + (64.10 - 46.80)^2 \end{aligned}}$$

$$D(1,1) = 34.43487$$

$$D(1,2) = \sqrt{\begin{aligned} &(62.80 - 253.90)^2 + (63.50 - 244.10)^2 \\ &+ (63.80 - 240.40)^2 \\ &+ (63.60 - 240.10)^2 + (64.10 - 244.50)^2 \end{aligned}}$$

$$D(1,2) = 404.9933$$

4. Proses perhitungan yang sama diulangi pada seluruh data di setiap provinsi pada setiap tahun mulai tahun 2017 hingga 2021.
5. Setiap data dikelompokkan dengan centroid yang terdekat. Setelah perhitungan K-Means selesai dilakukan, maka selanjutnya adalah mengelompokkan hasil perhitungan tersebut. Jika hasil *cluster* C1 lebih kecil dari C2 maka akan masuk ke *cluster* 1, begitu pula sebaliknya.
6. Menentukan posisi centroid baru dengan cara menghitung nilai rata-rata dari rata-rata dari data yang memilih pada centroid yang sama. Untuk menentukan centroid yang baru, dilakukan dengan mencari rata-rata dari nilai tiap-tiap daerah pada setiap centroid. Sebagai contoh, rumus untuk menentukan centroid pada *cluster* pertama yaitu:
 - a. Untuk *cluster* pertama (C1)

Pada iterasi ke-1 jumlah data yang masuk ke *cluster* 1 sebanyak 24 data, maka untuk menghitung pusat *cluster* baru adalah jumlah data dibagi dengan banyaknya data.

$$C1.1 = \frac{\sum \text{nilai hasil cluster 1 pada tahun 2017}}{\text{banyak cluster 1 pada tahun 2017}}$$

$$C1.1 = \frac{62.80+97.70+69.50+\dots+13.90}{24}$$

$$C1.1 = \frac{1338.5}{24}$$

$$C1.1 = 55.77083$$

Selanjutnya untuk tahun 2018

$$C1.2 = \frac{\Sigma \text{ nilai hasil cluster 1 pada tahun 2018}}{\text{banyak cluster 1 pada tahun 2018}}$$

$$C1.2 = \frac{63.50+99.40+78.00+\dots+15.20}{24}$$

$$C1.2 = \frac{1332.8}{24}$$

$$C1.2 = 55.53333$$

Selanjutnya dilakukan juga perhitungan sampai tahun 2021

b. Untuk *cluster* kedua (C2)

Pada iterasi ke-1 jumlah data yang masuk ke *cluster* 2 sebanyak 5 data, maka untuk menghitung pusat *cluster* baru adalah jumlah data dibagi dengan banyaknya data.

$$C2.1 = \frac{\Sigma \text{ nilai hasil cluster 2 pada tahun 2017}}{\text{banyak cluster 2 pada tahun 2017}}$$

$$C2.1 = \frac{390.60+158.80+253.90+260.70+187.40}{5}$$

$$C2.1 = \frac{1251.40}{5}$$

$$C2.1 = 250.28$$

Selanjutnya untuk tahun 2018

$$C2.2 = \frac{\Sigma \text{ nilai hasil cluster 2 pada tahun 2018}}{\text{banyak cluster 2 pada tahun 2018}}$$

$$C2.2 = \frac{392.70+172.60+244.10+262.50+193.90}{5}$$

$$C2.2 = \frac{1265.8}{5}$$

$$C2.2 = 253.16$$

Selanjutnya dilakukan juga perhitungan sampai tahun 2021

7. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama. Untuk proses selanjutnya mengulangi Langkah kedua, yaitu mencari jarak dari setiap data dengan menggunakan centroid baru. Proses ini terus diulangi hingga sudah tidak ada data lagi yang berpindah *cluster* atau centroid tidak berubah.

8. Menghitung Iterasi kedua

Untuk menghitung iterasi kedua, prosesnya sama seperti cara untuk menghitung iterasi pertama, hanya menyesuaikan pada bagian centroid. Untuk menghitung jarak antara data pertama terhadap tiap-tiap centroid adalah sebagai berikut:

$$D(1,1) = \sqrt{(62.80 - 55.77083)^2 + (63.50 - 55.53333)^2 + (63.80 - 55.35417)^2 + (63.60 - 54.70417)^2 + (64.10 - 55.2875)^2}$$

$$D(1,1) = 18.46632$$

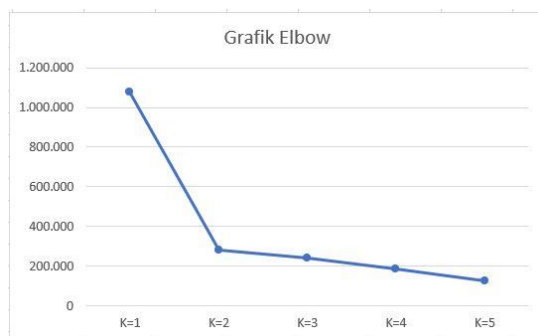
$$D(1,2) = \sqrt{(62.80 - 250,28)^2 + (63.50 - 253,16)^2 + (63.80 - 253,52)^2 + (63.60 - 250,76)^2 + (64.10 - 256,38)^2}$$

$$D(1,2) = 423.2184$$

Proses perhitungan yang sama diulangi pada seluruh data di setiap provinsi pada setiap tahun mulai tahun 2017 hingga 2021.

3. Hasil dan Pembahasan

Pada tahap penentuan jumlah *cluster* penelitian ini menggunakan metode Elbow. Pada perhitungan k=1 hingga k=5, maka hasilnya dapat dilihat pada Gambar 3.



Gambar 3. Hasil Metode Elbow

Berdasarkan grafik pada Gambar 3, terlihat bahwa bentuk elbow/siku terjadi pada kombinasi $k=2$. Berdasarkan hal ini maka data yang digunakan dalam penelitian ini akan lebih optimal jika dikelompokkan menjadi 2 cluster. Dan data ke-11 dan data ke-12 akan digunakan sebagai centroid awal dalam proses clustering. Centroid dengan nilai SSE minimum dapat dilihat pada Tabel 6. Berdasarkan hasil dari metode elbow didapatkan 2 centroid awal untuk menghitung proses clustering. Data centroid yang ditunjukkan pada Tabel 7

Tabel 6. Centroid dengan Nilai SSE Minimum

Centroid Ke = j	Data ke = i	Kelapa				
		2017	2018	2019	2020	2021
m1	11	53.40	47.40	48.10	46.50	46.80
m2	12	253.90	244.10	240.40	240.10	244.50

Tabel 7. Data Centroid

Data ke-	Centroid	Provinsi
11	C1	DI YOGYAKARTA
12	C2	JAWA TIMUR

Tabel 7 berisi dua provinsi yang akan digunakan sebagai centroid awal dalam proses clustering, hal ini sesuai dengan metode Elbow. Penentuan centroid hanya dilakukan satu kali, yaitu pada pada tanaman perkebunan kelapa karena komoditas tersebut adalah komoditas yang mengawali, dan komoditas-komoditas berikutnya menggunakan centroid awal yang sama dengan yang digunakan pada tanaman perkebunan kelapa. Hal ini dilakukan agar hasil penghitungan tidak bias dan tetap mempertahankan akurasi.

Suatu data akan menjadi anggota dari dari suatu cluster yang memiliki jarak terkecil dari pusat cluster-nya (centroid). Misalkan pada data pertama, jarak terkecil diperoleh pada cluster pertama, maka data tersebut akan masuk ke cluster 1. Demikian juga untuk data kedua dan selanjutnya. Untuk menentukan centroid yang baru, dilakukan dengan mencari rata-rata dari nilai tiap-tiap daerah pada setiap centroid. Setelah perhitungan selesai dilakukan, maka nilai yang dihasilkan adalah nilai yang digunakan sebagai centroid baru. Untuk hasil detailnya dapat dilihat pada Tabel 8.

Tabel 8. Centroid Baru untuk Iterasi 2

Centroid	2017	2018	2019	2020	2021
C1	55,77083	55,53333	55,35417	54,70417	55,2875
C2	250,28	253,16	253,52	250,76	256,38

Centroid baru pada Tabel 8 digunakan untuk perhitungan pada iterasi kedua. Hasil perhitungan terhadap setiap *cluster* pada iterasi 2 pada tanaman perkebunan kelapa dapat dilihat pada Tabel 9.

Tabel 9. Jarak Data pada Iterasi 2

Data ke	C1	C2
1	18,46632	423,2184
2	99,21955	342,4626
3	43,98587	398,9018
4	756,6165	315,1013
...
29	90,50774	532,1107

Pada tahap pengelompokan setiap data ke centroid terdekat, maka suatu data akan menjadi anggota dari suatu *cluster* yang memiliki jarak terkecil dari pusat *cluster*nya (centroid). Contoh hasil pengelompokan pada iterasi 2 pada tanaman perkebunan kelapa dapat dilihat pada tabel 10.

Tabel 10. Pengelompokan pada Iterasi 2

Data ke	C1	C2	Cluster
1	18,46632	423,2184	1
2	99,21955	342,4626	1
3	43,98587	398,9018	1
4	756,6165	315,1013	2
...
29	90,50774	532,1107	1

Misalkan pada data pertama, jarak terkecil diperoleh pada *cluster* pertama, maka data tersebut akan masuk ke *cluster* 1. Demikian juga untuk data kedua dan selanjutnya. Iterasi berhenti ketika kelompok data terakhir sama dengan kelompok data sebelumnya, dengan kata lain iterasi akan berhenti ketika centroid tidak berpindah tempat. Namun jika nilainya tidak sama, maka proses pengelompokan data diulangi kembali. Dari data di atas dapat disimpulkan bahwa iterasi berhenti pada iterasi 2. Dalam Penelitian ini menggunakan metode elbow untuk mencari *cluster* paling optimal. Untuk mendapatkan perbandingan *cluster* adalah dengan menghitung SSE (Sum of Square Error) dari masing masing *cluster* dengan perhitungan SSE sehingga mendapatkan *cluster* paling optimal $k=2$. Pengujian dilakukan dengan tanaman produksi tanaman perkebunan kelapa, kopi dan kakao. Untuk tiap-tiap dataset yang diuji, parameter yang digunakan untuk pengujian terhadap centroid *cluster* jumlah produksi sedikit (C1), dan banyak (C2). Hasil pengujian sistem terhadap data produksi tanaman perkebunan kelapa dengan jumlah record data yaitu 29 record yang dilakukan proses K-Means menghasilkan 2 *cluster*. Adapun Hasil Clustering Hasil Produksi Perkebunan ditunjukkan pada Tabel 11.

Tabel 11. Hasil Clustering

Tanaman Perkebunan	Provinsi (C1)	Provinsi (C2)
Kelapa	Aceh, Sumatera Utara, Sumatera Barat, Jambi, Sumatera Selatan, Bengkulu, Lampung, Jawa Barat, Di Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Papua	Riau, Jawa Tengah, Jawa Timur, Sulawesi Utara, Sulawesi Tengah
Kopi	Sumatera Barat, Riau, Jambi, Jawa Barat, Jawa Tengah, Di Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Papua	Aceh, Sumatera Utara, Sumatera Selatan, Bengkulu, Lampung, Jawa Timur
Kakao	Aceh, Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu, Jawa Barat, Jawa Tengah, Di Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Gorontalo, Maluku, Papua	Sumatera Barat, Lampung, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat

Berdasarkan Tabel 9 provinsi – provinsi yang masuk *cluster* 1 dengan jumlah produksi sedikit bisa dijadikan sebagai acuan pemerintah dalam meningkatkan hasil produksi kakao.

4. Penutup

Setelah mempelajari sejumlah permasalahan yang dihadapi dan sekaligus mencari solusi pemecahan permasalahan yang diajukan dapat diambil kesimpulan yaitu: Hasil *clustering* dari kelapa, kopi, dan kakao menghasilkan 2 *cluster* yang diperoleh dari perhitungan Sum of Squared Error (SSE) dengan nilai minimum 279261.63 yaitu produksi sedikit, dan produksi banyak. Hasil *clustering* kelapa diperoleh *cluster* produksi sedikit sebanyak 24 provinsi, *cluster* banyak sebanyak 5 provinsi, untuk kopi diperoleh *cluster* produksi sedikit sebanyak 23 provinsi, *cluster* banyak sebanyak 6 provinsi, dan kakao diperoleh *cluster* produksi sedikit sebanyak 23 provinsi, *cluster* banyak sebanyak 6 provinsi. Hasil *clustering* dari kelapa, kopi, dan kakao yang termasuk dalam *cluster* 1 merupakan provinsi - provinsi yang memiliki hasil produksi sedikit sehingga dapat dijadikan sebagai acuan pemerintah untuk meningkatkan hasil produksi kelapa, kopi, dan kakao.

5. Referensi

- [1] J. I. Ekonomi, "Analisis ekspor kakao indonesia di pasar internasional," vol. 1, no. Machfudz 2007, pp. 453–472, 2017.
- [2] J. Ilmiah and W. Pendidikan, "Pengaruh Ekspor Dan Impor Terhadap Cadangan Devisa Tahun 2018-2022," vol. 9, no. 12, pp. 571–579, 2023.
- [3] E. Y. Rahmawati, B. Ismanto, D. S. Sitorus, and I. Artikel, "Analisis Pengaruh Ekspor Dan Kurs Terhadap Cadangan Devisa Indonesia Tahun 1990-2019," no. 2016, 2019.
- [4] A. BASTIAN, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *J. Sist. Inf.*, vol. 14, no. 1, pp. 28–34, 2018, doi: 10.21609/jsi.v14i1.566.
- [5] A. K. Wardhani, "K-Means Algorithm Implementation for Clustering of Patients Disease in

- Kajen Clinic of Pekalongan,” *J. Transform.*, vol. 14, no. 1, p. 30, 2016, doi: 10.26623/transformatika.v14i1.387.
- [6] E. Muningsih and S. Kiswati, “Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan *Clustering* Pelanggan,” *Joutica*, vol. 3, no. 1, p. 117, 2018, doi: 10.30736/jti.v3i1.196.
- [7] P. Meilina, “Penerapan Data Mining dengan Metode Klasifikasi Menggunakan Decision Tree dan Regresi,” *J. Teknol. Univ. Muhammadiyah Jakarta*, vol. 7, no. 1, pp. 11–20, 2015, [Online]. Available: jurnal.ftumj.ac.id/index.php/jurtek.
- [8] L. Maulida, “Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, p. 167, 2018, doi: 10.14421/jiska.2018.23-06.
- [9] D. A. Alodia, A. P. Fialine, D. Endriani, and E. Widodo, “Implementasi Metode K-Medoids *Clustering* untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan,” *Sepren*, vol. 2, no. 2, pp. 1–13, 2021, doi: 10.36655/sepren.v2i2.606.
- [10] Z. Mustofa and Iman Saufik Suasana, “Algoritma *Clustering* K-Medoids Pada E-Government Bidang Information and Communication Technology Dalam Penentuan Status Edgi,” *J. Teknol. Inf. Dan Komun.*, vol. 9, no. 1, pp. 1–10, 2020, doi: 10.51903/jtikp.v9i1.162.
- [11] D. Alfatah, “Application of the K-Means *Clustering* Algorithm in Mapping the Regional Voter Strategy for the Legislative Candidates for the DPR RI Penerapan Algoritma K-Means *Clustering* dalam Memetakan Strategi Daerah Pemilih pada Calon Legislatif DPR RI,” *J. Kom.*, vol. 1, no. 2, pp. 435–443, 2021.
- [12] E. Muningsih, I. Maryani, and V. R. Handayani, “Penerapan Metode K-Means dan Optimasi Jumlah *Cluster* dengan Index Davies Bouldin untuk *Clustering* Propinsi Berdasarkan Potensi Desa,” *J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839>.
- [13] S. Y. Riska and L. Farokhah, “Perbandingan Hasil Evaluasi Algoritma K-Means dan K-Medoid Berdasarkan Kunjungan Wisatawan Mancanegara ke Indonesia,” *J. Inf. Technol.*, vol. 8, no. 1, p. hal. 1-8, 2023.
- [14] N. T. Hartanti, “Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.
- [15] A. Winarta and W. J. Kurniawan, “Optimasi *Cluster* K-means Menggunakan Metode Elbow pada Data Pengguna Narkoba dengan Pemrograman Python,” *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.
- [16] A. Aditya, I. Jovian, and B. N. Sari, “Implementasi K-Means *Clustering* Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019,” *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.