

Analisis Cluster Luas Lahan Panen Tanaman Biofarma di Indonesia Tahun 2022

Irfan Zulkarnaen Anas^{1*}, Rahmadi Yotenka¹

¹ Program Studi Statistika, Universitas Islam Indonesia, Jl.Kaliurang KM 14,5, Kabupaten Sleman Daerah Istimewa Yogyakarta, 55584, Indonesia

*Corresponding author: 21611002@students.uii.ac.id



P-ISSN: 2986-4178
E-ISSN: 2988-4004

Riwayat Artikel

Dikirim: 14 November 2024
Direvisi: 14 Januari 2025
Diterima: 20 Januari 2025

ABSTRAK

Tanaman biofarmaka adalah tanaman yang digunakan sebagai bahan dasar untuk obat-obatan, kosmetik, dan produk kesehatan lainnya. Tanaman ini menghasilkan simplisia, yaitu bahan alamiah yang dipergunakan sebagai obat yang belum mengalami pengolahan apapun, biasanya berupa bagian tanaman seperti akar, daun, batang, atau eksudat tanaman (zat yang keluar dari tanaman). Jenis-jenis tanaman biofarmaka antara lain jahe, laos/lengkuas, kencur, kunyit, lempuyangan, temulawak, temuireng, kapulaga, mengkudu/pace, makota dewa, kejibeling, sambiroto, dan lidah buaya. Tujuan dari penelitian ini adalah untuk mengetahui perbedaan jumlah cluster dengan menggunakan metode *elbow* dan *silhouette*, dengan menggunakan algoritma *k-means* untuk analisis clusternya, dari data luas lahan panen tanaman biofarmaka tahun 2022 dari BPS. dari hasil analisis diperoleh untuk metode *elbow* optimal dengan 3 cluster sedangkan untuk metode *silhouette* optimal dengan 2 cluster. Untuk hasil dari metode *elbow* pada cluster pertama terdiri dari 31 provinsi dan untuk anggota cluster kedua yaitu Jawa Timur, dan untuk anggota cluster ke tiga yaitu Jawa barat dan Jawa Tengah. Dan untuk hasil metode *silhouette* pada cluster pertama terdiri dari 31 provinsi dan untuk anggota cluster kedua terdapat 3 provinsi.

Kata Kunci: biofarmaka, *clustering*, *elbow*, *silhouette*, *k-means*.

ABSTRACT

Biopharmaceutical plants are plants that are used as basic ingredients for medicines, cosmetics, and other health products. These plants produce simplisia, which is natural material used as medicine that has not undergone any processing, usually in the form of plant parts such as roots, leaves, stems, or plant exudates (substances that come out of plants). Types of biopharmaca plants include ginger, galangal, east Indian galangal, turmeric, zingiber aromaticum, java turmeric, black turmeric, cardamom, Indian mulberry, phaleria macrocarpa, verbenaceae, king of bitter, and aloe vera. The purpose of this study is to determine the difference in the number of clusters using the elbow and silhouette methods, using the k-means algorithm for cluster analysis, from data on the area of harvested biopharmaca plants in 2022 from BPS. from the analysis results obtained for the optimal elbow method with 3 clusters while for the optimal silhouette method with 2 clusters. For the results of the elbow method, the first cluster consists of 31 provinces and for the second cluster members, namely East Java, and for the third cluster members, namely West Java and Central Java. And for the results of the silhouette method in the first cluster consists of 31 provinces and for the second cluster members there are 3 provinces.

Keywords: *biopharmaca, clustering, elbow, silhouette, k-means.*

1. Pendahuluan

Tanaman biofarmaka adalah tanaman yang digunakan sebagai bahan dasar untuk obat-obatan, kosmetik, dan produk kesehatan lainnya. Tanaman ini menghasilkan simplisia, yaitu bahan alamiah yang dipergunakan sebagai obat yang belum mengalami pengolahan apapun, biasanya berupa bagian tanaman seperti akar, daun, batang, atau eksudat tanaman (zat yang keluar dari tanaman). Jenis-jenis tanaman biofarmaka antara lain jahe, laos/lengkuas, kencur, kunyit, lempuyangan, temulawak, temuireng, kapulaga, mengkudu/pace, makota dewa, kejobeling, sambiroto, dan lidah buaya. Tanaman biofarmaka memiliki berbagai manfaat yang penting untuk kesehatan seperti kunyit berkhasiat sebagai anti oksidan, anti mikroba, pencegah kanker, antitumor, serta dapat menurunkan kadar lemak darah dan kolesterol, dan sebagai pembersih darah, ada tanaman mengkudu yang bermanfaat untuk meningkatkan sistem kekebalan tubuh, menurunkan tekanan darah, dan memiliki sifat antikanker serta antioksidan, dan ada juga jahe untuk mengatasi masalah pencernaan, mual, dan masuk angin. Selain ketiga tanaman tersebut, masih banyak tanaman biofarmaka lainnya yang memberikan manfaat besar bagi kesehatan, seperti kapulaga, lidah buaya, dan temulawak, yang semuanya memiliki kegunaan yang unik dan berharga dalam menjaga dan meningkatkan kesehatan manusia [1].

Namun, di Indonesia, meskipun banyak lahan yang digunakan untuk menanam tanaman biofarmaka, pengelolaan lahan yang belum optimal menyebabkan hasil panen yang didapatkan belum maksimal di beberapa provinsi [2]. Oleh karena itu, diperlukan upaya lebih lanjut untuk mengatasi tantangan ini, salah satunya melalui pengelompokan provinsi di Indonesia berdasarkan potensi dan kondisi masing-masing [2]. Dengan

pengelompokan ini, pemerintah dapat menerapkan strategi yang lebih tepat dan terfokus untuk mengembangkan produksi tanaman biofarmaka di setiap daerah, memaksimalkan potensi lahan yang ada, dan akhirnya meningkatkan kesejahteraan petani serta ketersediaan bahan baku berkualitas tinggi untuk industri obat-obatan, kosmetik, dan produk kesehatan lainnya. Untuk metode pengelompokannya sendiri menggunakan metode analisis *k-means* [2]. *K-means* adalah algoritma pengelompokan data (*clustering*) yang berfungsi untuk membagi sekumpulan data menjadi beberapa kelompok atau cluster berdasarkan kesamaan karakteristiknya. Prosesnya bersifat partisi, artinya data akan dipecah menjadi sejumlah cluster yang telah ditentukan sebelumnya [2]. Metode ini dipilih karena memiliki algoritma yang secara iteratif memperbarui pusat cluster hingga mencapai stabilitas, memastikan bahwa cluster yang dihasilkan adalah yang paling representatif dari data. Untuk optimalisasi jumlah cluster menggunakan metode *elbow* dan *silhouette* [2]. Metode *elbow* adalah teknik untuk menentukan jumlah cluster yang optimal dengan cara mengukur jumlah total variansi (dalam *k-means*, ini biasanya dihitung sebagai jumlah total dari kuadrat jarak antara titik data dan pusat cluster terdekat, disebut sebagai SSE (*Sum of Squared Errors*) untuk berbagai nilai *k* [3]. metode *silhouette* adalah teknik lain yang digunakan untuk menentukan jumlah cluster yang optimal dengan cara mengevaluasi seberapa baik data dikelompokkan. Kedua metode ini dipilih karena mereka berdua memberikan perspektif yang berbeda. *Elbow Method* fokus pada penurunan SSE, sedangkan *Silhouette Method* fokus pada kualitas clusterisasi dari sudut pandang jarak antar-cluster dan dalam-cluster [3]. Menggunakan kedua metode ini bersama-sama memberikan evaluasi yang lebih komprehensif. Keduanya saling melengkapi, memberikan baik evaluasi kuantitatif maupun kualitatif tentang seberapa baik data dikelompokkan, dan bersama-sama, mereka membantu memastikan bahwa clusterisasi yang dilakukan adalah representatif dan bermakna [2].

Penelitian yang dilakukan oleh Arif, Rizki dan Mulyawan pada tahun 2023 dengan judul "Perbandingan algoritma *k-means* dengan *k-medoids* dalam pengelompokan kabupaten dan kota berdasarkan tanaman biofarmaka". Penelitian ini dilakukan di Jawa Barat dan didapatkan hasil untuk metode *k-means* diperoleh 4 cluster dan untuk metode *k-medoids* juga menghasilkan 4 cluster. Untuk membandingkan kedua metode dilakukan perbandingan menggunakan *cluster variance* didapatkan hasil nilai DBI *k-means* sebesar 0.087 sedangkan *k-medoids* sebesar 0.428. Hal ini menunjukkan bahwa pada pengujian DBI *k-means* lebih rendah dibandingkan dengan *k-medoids* karena semakin rendah nilai yang diperoleh maka kualitas cluster yang dihasilkan semakin baik.

Kedua penelitian yang dilakukan oleh Guntara dan Lutfi pada tahun 2023 dengan judul "Cacah cluster pada clusterisasi dengan algoritma *k-means* menggunakan *silhouette coefficient* dan *elbow method*". Penelitian ini menggunakan data mahasiswa *dropout* (DO) dengan variabel yang digunakan NIM, jumlah semester tidak aktif, jumlah sks, dan IPK dengan 955 mahasiswa. Didapatkan hasil bahwa dengan menggunakan metode *elbow* ataupun *silhouette* sama-sama menunjukkan bahwa jumlah cluster optimal adalah 2 klastre dengan cluster 1 terdata 813 mahasiswa dan cluster 2 terdapat 142 mahasiswa. Pada cluster 1 didominasi mahasiswa DO karena $IPK < 2.00$, jumlah sks yang diambil < 90 sks, dan tidak aktif kuliah ($cuti \geq 4$ semester).

Penelitian ini bertujuan untuk mengetahui perbedaan jumlah cluster dengan menggunakan metode *elbow* dan *silhouette*, dengan algoritma *k-means* dari data luas lahan panen tanaman biofarmaka dari seluruh provinsi di Indonesia tahun 2022.

2. Metodologi Penelitian

Penelitian ini menggunakan data sekunder dari Badan Pusat Statistik (BPS) tentang luas lahan panen tanaman biofarmaka pada tahun 2022 di 34 provinsi di Indonesia.

Penelitian ini melihat tujuh jenis tanaman biofarmaka yaitu jahe, kapulaga, kencur, lidah buaya, mahkota dewa, jeruk nipis, dan kunyit. Penelitian ini juga menggunakan metode analisis cluster non-hirarki menggunakan algoritma *k-means* dan metode *elbow* dan *silhouette* untuk menghitung jumlah cluster.

Clusterisasi, juga dikenal sebagai *clustering*, adalah teknik untuk menempatkan data ke dalam berbagai kelompok (*cluster*). Di dalam satu kelompok, objek memiliki banyak kesamaan dan perbedaan yang signifikan satu sama lain. Kesamaan dan perbedaan ini biasanya didasarkan pada nilai atribut atau jarak antara objek. Untuk clusterisasi data, ada dua pendekatan utama yaitu pendekatan partisi (non-hirarki) dan pendekatan hirarki. Pendekatan pertama menggunakan pemisahan data yang telah dianalisis ke dalam *cluster* yang sudah ada, dimulai dengan menentukan jumlah *cluster* yang diinginkan. Clusterisasi dilakukan tanpa mengikuti proses hirarki setelah jumlah *cluster* telah ditentukan. *Cluster K-Means* adalah salah satu metode yang paling umum digunakan dalam pendekatan ini untuk mengelompokkan banyak objek. Metode ini sangat efektif dan efisien untuk mengelompokkan banyak objek. Sebaliknya, pendekatan hirarki mengelompokkan data dengan menggabungkan dua atau lebih objek yang memiliki kesamaan paling dekat. Selanjutnya, objek lain digabungkan hingga terbentuk struktur pohon yang jelas (*dendrogram*), yang menunjukkan tingkatan hirarki dari yang paling mirip hingga yang paling tidak mirip [3].

2.1 K-Means

Algoritma *k-means* adalah algoritma partisi yang menentukan jumlah awal *cluster* dengan menetapkan nilai *centroid* awal. Proses ini dilakukan berulang-ulang untuk membentuk basis data *cluster* dan membutuhkan jumlah *cluster* awal sebagai input dan menghasilkan *cluster* akhir sebagai output. Jika algoritma dirancang untuk menghasilkan *k cluster*, maka akan ada *k cluster* awal dan *k cluster* akhir. Metode *k-means* menggunakan pola *k* sebagai titik awal *centroid* secara acak. Posisi awal *centroid* yang dipilih secara acak memengaruhi jumlah iterasi yang diperlukan untuk mencapai *cluster centroid*, dan iterasi berlanjut hingga posisi *centroid* tidak berubah. Rumus jarak digunakan untuk menghitung nilai *k* yang dipilih sebagai pusat awal. Rumus untuk mencari jarak terdekat antara *centroid* dan data atau objek. Data yang berada di dekat *centroid* akan dimasukkan ke dalam *cluster* yang sama [3].

K-means clustering menggunakan parameter *input k* dan mengelompokkan sekumpulan *n* objek ke dalam *k cluster*. Tujuan utamanya adalah memastikan tingkat kemiripan antara anggota dalam satu cluster menjadi tinggi, sementara tingkat kemiripan dengan anggota cluster lainnya menjadi rendah. Untuk mengukur kemiripan anggota dengan cluster, digunakan konsep kedekatan objek terhadap nilai rata-rata (*mean*) pada *centroid* cluster [4]. Berikut ini adalah langkah-langkah dalam algoritma *k-means*.

1. Menentukan jumlah *cluster* (*k*) pada *dataset*
2. Menentukan *centroid*, penentuan nilai *centroid* pada tahap awal ditentukan secara random, sedangkan pada tahap iterasi digunakan rumur seperti pada persamaan berikut ini:

$$V_k = \frac{\sum_{i=1}^{N_i} X_i}{N_k} \quad (1)$$

dengan:

V_k = Centeroid pada *cluster* ke *k*

X_i = Objek ke *i*

N_k = Banyaknya objek/jumlah data yang menjadi anggota *cluster* ke *k*

- Menghitung jarak terdekat *centroid*. Jarak *centroid* yang digunakan adalah *Euclidean Distance*.

$$D_E = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, i = 1, 2, 3, \dots, n \quad (2)$$

dengan:

D_E = *Euclidean Distance*

n = Banyaknya Objek

x_i = Objek x ke- i

y_i = Objek y ke- i

- Setelah itu alokasikan setiap data ke dalam *cluster* yang mempunyai jarak terdekat.
- Lakukan kembali langkah kedua dan ketiga sampai *centroid* stabil.

Metode *Elbow* adalah metode untuk melihat persentase perubahan jumlah *cluster* yang membentuk siku pada suatu titik dalam grafik untuk menentukan jumlah *cluster* terbaik. Ini dimulai dengan nilai *cluster* awal dan kemudian ditambahkan ke model data untuk menentukan *cluster* terbaik. Jumlah *cluster* yang ditambahkan dibandingkan dengan persentase hasil perhitungan. Grafik menunjukkan perbedaan persentase dari setiap nilai *cluster*. Jumlah *cluster* dianggap optimal jika perbandingan antara *cluster* pertama dan kedua menunjukkan sudut tajam atau penurunan terbesar dalam grafik [5].

Metode *silhouette* menggabungkan kohesi dan separasi untuk menilai *cluster*. Menghitung jumlah objek dalam setiap *cluster* menunjukkan kohesi, sedangkan menghitung jarak rata-rata antara setiap objek dalam setiap *cluster* ke *cluster* terdekat menunjukkan separasi. Untuk mengetahui jarak antar data, rumus jarak geometris digunakan. Nilai *silhouette* untuk seluruh data dengan jumlah *cluster* k disebut $sil(k)$, dan nilai *silhouette* untuk setiap *cluster* atau seluruh *cluster* dari hasil algoritma *clustering* dapat dihitung untuk mengevaluasi kualitas hasil *clustering*. $Sil(k)$ adalah nilai *silhouette* rata-rata untuk semua *cluster* [5].

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^k sil(c_i) \quad (3)$$

dengan:

$sil(c)$: rata-rata nilai *silhouette*

$sil(k)$: nilai *silhouette* semua *cluster*

$|k|$: banyaknya *cluster* k

2.2 Multikolinieritas

Uji multikolinieritas merupakan metode untuk menilai adanya korelasi yang signifikan antar variabel independen dalam model regresi linier berganda. Korelasi yang kuat antar variabel independen dapat mengganggu hubungan antara variabel independen dan dependen. Uji multikolinieritas yang dilakukan tidak hanya berfokus pada derajat korelasi antar variabel independen, tetapi juga pada nilai toleransi dan VIF (*Variance Inflation Factor*). Berikut adalah rumus untuk menghitung nilai VIF:

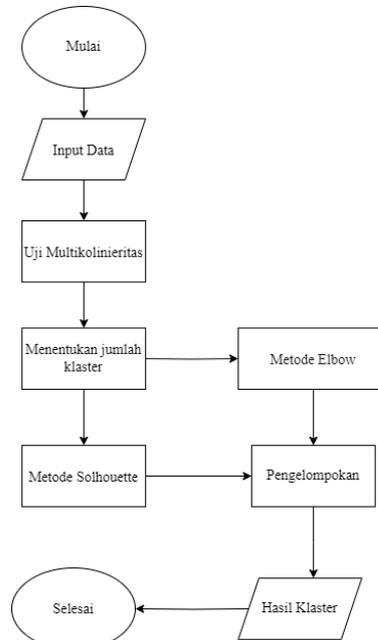
$$VIF = \frac{1}{1 - R_i^2} \quad (4)$$

Untuk menyimpulkan bahwa suatu model regresi bebas dari multikolinieritas, nilai VIF seharusnya tidak melebihi 10, dan nilai *tolerance* seharusnya tidak kurang dari 0.10.

Dengan kata lain, jika model regresi memiliki nilai VIF yang rendah dan nilai *tolerance* yang tinggi, dapat dikatakan bahwa model tersebut bebas dari multikolinieritas [6].

2.3 Diagram Alir

Berikut adalah tahapan penelitian yang dilakukan.



Gambar 2.1 Diagram alir

1. Melakukan input data yang digunakan, yaitu data luas lahan panen tanaman biofarmaka di Indonesia pada tahun 2022.
2. Selanjutnya dilakukan uji multikolinieritas pada variabel independen.
3. Tentukan jumlah *cluster* yang optimal untuk digunakan pada algoritma *k-means*. Dengan menggunakan metode *elbow* dan *silhouette*.
4. Masukkan hasil jumlah *cluster* optimal kedalam algoritma *k-means* sebagai nilai *k*.
5. Selanjutnya didapatkan hasil clusterisasi dan selesai.

3. Hasil dan Pembahasan

Pada bab ini akan dibahas analisis deskriptif dari data, uji multikolinieritas dan hasil analisis *clustering*.

3.1. Analisis Deskriptif

Penelitian ini dilakukan dengan cara mengambil data luas lahan panen tanaman biofarmaka di Indonesia tahun 2022 dari *website* bps.go.id. Data yang akan diuji adalah data luas lahan panen jahe, kapulaga, kencur, kunyit, lidah buaya, mahkota dewa, dan jeruk nipis pada tahun 2022 di seluruh provinsi di Indonesia.

Tabel 3.1 Statistika deskriptif

	Jahe	Kapulaga	Kencur	Kunyit	Lidahbuaya	Mahkota	Jeruk Nipis
Mean	3054408	1816094	921781	2222256	36435	6096	65316
Max	20813926	37330964	11412697	40738006	814183	69765	566720
Min	743	0	331	449	0	0	0
Median	923342	2434	103725	451405	1533	898	37959

Dari Tabel 3.1 diperoleh hasil bahawa rata-rata luas lahan panen jahe, kapulaga, kencur, kunyit, lidah buaya, mahkota dewa, dan jeruk nipis secara berturut-turut sebesar 3054408m², 1816094m², 921781m², 2222256m², 36345m², 6096m², dan 65316m².



Gambar 3.1 Luas Lahan Panen Jahe Tahun 2022

Gambar 3.1 merupakan grafik persebaran luas lahan panen tanaman jahe di Indonesia tahun 2022. Dari grafik tersebut dapat dilihat bahwa provinsi Jawa Timur memiliki luas lahan panen jahe paling tinggi di Indonesia dengan luas lahan panen sebesar 20813926m² sedang DKI Jakarta menjadi provinsi dengan luas lahan panen jahe paling kecil di Indonesia sebesar 743m², dan untuk rata-rata luas lahan panen tanaman jahe di Indonesia sebesar 3054408m².



Gambar 3.2 Luas Lahan Panen Kapulaga Tahun 2022

Gambar 3.2 merupakan grafik persebaran luas lahan panen tanaman kapulaga di Indonesia tahun 2022. Dari grafik tersebut dapat dilihat bahwa provinsi Jawa Barat memiliki luas lahan panen kapulaga paling tinggi di Indonesia dengan luas lahan panen sebesar 37330964m² sedang ada beberapa wilyah seperti Kep Riau, DKI Jakarta, Bali, Kalimantan Selatan, Gorontalo, Maluku, dan Maluku Utara yang menjadi provinsi dengan luas lahan panen kapulaga paling kecil di Indonesia sebesar 0m² atau tidak memiliki lahan panen, dan untuk rata-rata luas lahan panen tanaman kapulaga di Indonesia sebesar 1816094m².



Gambar 3.3 Luas Lahan Panen Kencur Tahun 2022

Gambar 3.3 merupakan grafik persebaran luas lahan panen tanaman kencur di Indonesia tahun 2022. Dari grafik tersebut dapat dilihat bahwa provinsi Jawa Tengah memiliki luas lahan panen kencur paling tinggi di Indonesia dengan luas lahan panen sebesar 11412697m² sedang DKI Jakarta menjadi provinsi dengan luas lahan panen kencur paling kecil di Indonesia sebesar 331m², dan untuk rata-rata luas lahan panen tanaman kencur di Indonesia sebesar 921781m².



Gambar 3.4 Luas Lahan Panen Kunyit Tahun 2022

Gambar 3.4 merupakan grafik persebaran luas lahan panen tanaman kunyit di Indonesia tahun 2022. Dari grafik tersebut dapat dilihat bahwa provinsi Jawa Timur memiliki luas lahan panen kunyit paling tinggi di Indonesia dengan luas lahan panen sebesar 40738006m² sedang DKI Jakarta menjadi provinsi dengan luas lahan panen kunyit paling kecil di Indonesia sebesar 449m², dan untuk rata-rata luas lahan panen tanaman kunyit di Indonesia sebesar 2222256m².



Gambar 3.5 Luas Lahan Panen Lidah Buaya Tahun 2022

3.2. Pengujian Multikolinieritas

Pengujian multikolinieritas digunakan untuk menentukan apakah ada atau tidak korelasi antar variabel. Nilai faktor variasi inflasi (VIF) dapat digunakan untuk melakukan pemeriksaan ini. Jika nilai VIF lebih besar dari 10, maka terdapat multikolinieritas, atau korelasi antar variabel. Nilai VIF yang dihasilkan dapat dilihat pada tabel 3.2 dibawah ini.

Tabel 3.2 Nilai VIF variabel

Variabel	VIF	Keputusan
Kunyit	1.14265	Gagal tolak H_0
Jahe	3.30229	Gagal tolak H_0
Kapulaga	2.05379	Gagal tolak H_0
Kencur	2.19177	Gagal tolak H_0
Lidah buaya	1.28879	Gagal tolak H_0
Jeruk nipis	4.45500	Gagal tolak H_0
Mahkota dewa	6.85802	Gagal tolak H_0

Pengujian Hipotesis sebagai berikut:

I. Uji hipotesis

H_0 : tidak terjadi multikolinieritas

H_1 : terjadi multikolinieritas

II. Daerah Kritis: $VIF > 10$ maka tolak H_0

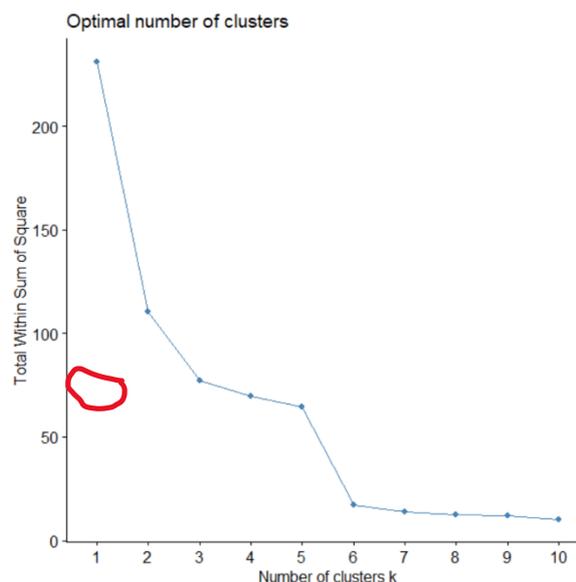
III. Keputusan

Gagal tolak H_0 karena $VIF < 10$

IV. Kesimpulan

Dengan data yang digunakan dapat disimpulkan bahwa gagal tolak H_0 dan menyatakan bahwa tidak ada korelasi atau hubungan antar variabel independen atau tidak terjadi multikolinieritas.

3.2.1. Metode Elbow

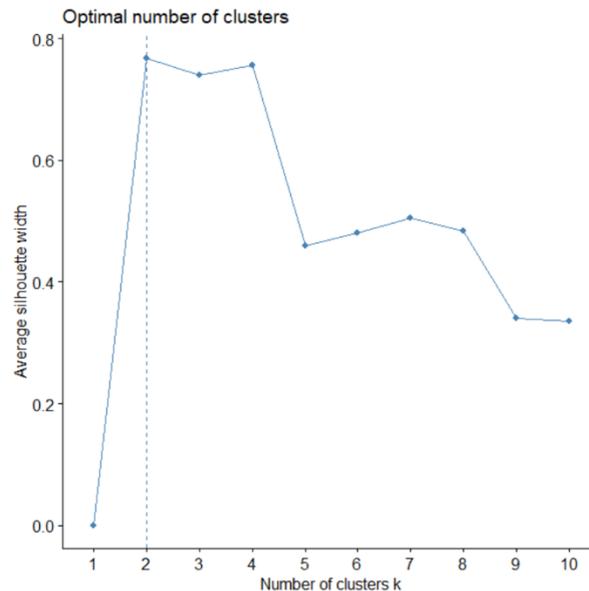


Gambar 3.8 Grafik *Elbow*

Metode *elbow* sebagai penentu k optimal adalah salah satu pendekatan yang paling umum untuk menentukan k optimal. Gambar 3.8 menunjukkan bahwa garis mengalami patahan

yang membentuk siku atau *elbow* pada saat $k = 2$. Dengan demikian, K optimal diperoleh pada saat $k = 2$.

3.2.2. Metode Silhouette



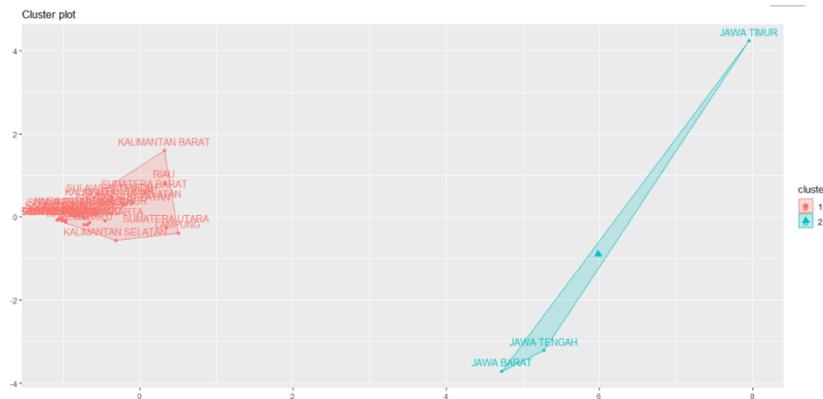
Gambar 3.9 Grafik metode *Silhautte*

Pendekatan nilai rata-rata *silhouette* digunakan untuk mengukur kualitas *cluster* yang terbentuk. Kualitas *cluster* yang lebih baik memiliki nilai rata-rata yang lebih tinggi. Jumlah *cluster* yang ideal terbentuk pada $k = 2$, seperti yang ditunjukkan oleh grafik pada Gambar 3.9.

Setelah dilakukan pendekatan jumlah *cluster* yang optimal untuk digunakan didapatkan hasil bahwa 2 adalah jumlah *cluster* yang optimal baik menggunakan metode *elbow* atau metode *silhouette*. Dan berikut adalah hasil clusterisasinya

Tabel 3.3 Hasil Clusterisasi Berdasarkan Provinsi

<i>Cluster</i>	Anggota
1	Aceh, Riau, Jambi, Bengkulu, Kep Bangka Belitung, Kep Riau, Sumatera Utara, Sumatera Barat, Sumatera Selatan, Lampung, Banten, DKI Jakarta, DI Yogyakarta, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Utara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Papua Barat, Papua
2	Jawa Timur, Jawa Barat, Jawa Tengah



Gambar 3.10 Visualisai Hasil Cluster

Dari Tabel 3.3 dan Gambar 3.10 didapatkan hasil bahwa luas lahan panen jahe, kapulaga, kencur, kunyit, lidah buaya, mahkota dewa, dan jeruk nipis diperoleh hasil pengelompokan sebanyak 2 kelompok dengan menggunakan *cluster k-means*, pada *cluster* kedua terdapat provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat, dan pada *cluster* pertama terdapat 31 provinsi.

Tabel 3.4 Profilisasi Cluster Luas Lahan Panen Dengan Satuan Meter Persegi

Cluster	Kunyit	Jahe	Kapulaga	Kencur	Lidah buaya	Mahkota dewa	Jeruk nipis
1	586217.2	1411819	79831	398022.2	34092.58	3150.94	43651.13
2	18127994.3	20027834	19757470	6333958.3	60637.33	36530.33	29188.33

Berdasarkan Tabel 3.4, maka dapat dilakukan profilisasi tiap kelompok yang terbentuk. Pada *cluster* 1 rata-rata luas lahan panen untuk tanaman kunyit sebesar 586217.2m², rata-rata luas lahan panen untuk tanaman jahe sebesar 1411819m², rata-rata luas lahan panen untuk tanaman kapulaga sebesar 79831m², rata-rata luas lahan panen untuk tanaman kencur sebesar 398022.2m², rata-rata luas lahan panen untuk tanaman lidah buaya sebesar 34092.58m², rata-rata luas lahan panen untuk tanaman mahkota dewa sebesar 3150.94m², dan rata-rata luas lahan panen untuk tanaman jeruk nipis sebesar 43651.13m². Pada *cluster* 2 rata-rata luas lahan panen untuk tanaman kunyit sebesar 18127994.3m², rata-rata luas lahan panen untuk tanaman jahe sebesar 20027834m², rata-rata luas lahan panen untuk tanaman kapulaga sebesar 19757470m², rata-rata luas lahan panen untuk tanaman kencur sebesar 6333958.3m², rata-rata luas lahan panen untuk tanaman lidah buaya sebesar 60637.33m², rata-rata luas lahan panen untuk tanaman mahkota dewa sebesar 36530.33m², dan rata-rata luas lahan panen untuk tanaman jeruk nipis sebesar 29188.33m².



Gambar 3.13 Visualisai Hasil Cluster

Gambar 3.13 adalah peta penyebaran luas lahan panen tanaman biofarmaka di Indonesia tahun 2022 yang terdiri dari 2 cluster yaitu cluster 1 berwarna biru dan cluster 2 berwarna merah. Berdasarkan hasil yang didapatkan bahwa persebaran luas lahan panen tanaman biofarmaka tahun 2022 lebih banyak tersebar di pulau Jawa. Cluster 2 memiliki nilai rata-rata lebih tinggi daripada cluster 1 yang dimana cluster 2 di isi oleh 3 provinsi di pulau Jawa dapat diartikan bahwa 3 provinsi di pulau Jawa memiliki luas lahan panen tanaman biofarmaka lebih besar di bandingkan 31 provinsi lainnya.

4. Kesimpulan

Dari data luas lahan panen jahe, kapulaga, kencur, kunyit, lidah buaya, mahkota dewa, dan jeruk nipis diperoleh hasil pengelompokan sebanyak 2 kelompok dengan menggunakan *cluster k-means*, pada cluster kedua terdapat provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat, dan pada cluster pertama terdapat 31 provinsi. Cluster 2 memiliki nilai rata-rata lebih tinggi daripada cluster 1 yang dimana cluster 2 di isi oleh 3 provinsi di pulau Jawa dapat diartikan bahwa 3 provinsi di pulau Jawa memiliki luas lahan panen tanaman biofarmaka lebih besar di bandingkan 31 provinsi lainnya.

5. Daftar Pustaka

- [1] B. P. Statistik, "bps.go.id," 13 februari 2023. [Online]. Available: <https://www.bps.go.id/indicator/55/63/1/produksi-tanaman-biofarmaka-obat-.html>.
- [2] M. Guntara and N. Lutfi, "Cacah Klaster pada Klasterisasi Dengan Algoritma K-Means Menggunakan Silhouette Coeficient dan Metode Elbow," *JuTI"Jurnal Teknologi Informasi"*, vol. 2, no. 1, 2023.
- [3] F. M. Izzadin, "Optimasi Jumlah Klaster Dengan K-Means Dengan Metode Elbow dan Silhouette," [dspace.uui.ac.id](https://dspace.uui.ac.id/bitstream/handle/123456789/28446/16611050%20Fata%20Mukhammad%20Izzadin.pdf?sequence=1&isAllowed=y), 2020. [Online]. Available: <https://dspace.uui.ac.id/bitstream/handle/123456789/28446/16611050%20Fata%20Mukhammad%20Izzadin.pdf?sequence=1&isAllowed=y>.
- [4] T. Khotimah, "Pengelompokan Surat Dalam Al-Qur'an Menggunakan Algoritma K-Means," *SIMETRIS*, pp. 83-88, 2014.
- [5] A. M. Arif, "Perbandingan Algoritma K-Means dan K-Medoids Dalam Pengelompokan Kabupaten dan Kota Berdasarkan Tanaman Biofarmaka," *JATI(jurnal amhasiswa teknik informatika)*, vol. 7, 2023.
- [6] Setiawati, "Analisis Pengaruh Kebijakan Deviden Terhadap Nilai Perusahaan Pada Perusahaan Farmasi di BEI," *jurnal inovasi penelitian*, vol. 1, no. 8, 2021.