

## Implementasi Algoritma *K-Means* dalam Analisis *Cluster* Usaha Pertanian Perorangan di Provinsi D. I Yogyakarta Tahun 2023

Faza Izzatul Muttaqin<sup>1\*</sup>, Zahra Rizky Ramadani<sup>2</sup>, Rahmadi Yotenka<sup>3</sup>, Suparna<sup>4</sup>

<sup>1,2,3</sup> Program Studi Statistika, Universitas Islam Indonesia, Jl. Kaliurang KM 14,5, Kabupaten Sleman Daerah Istimewa Yogyakarta, 55584, Indonesia

<sup>4</sup> Badan Pusat Statistik (BPS) Provinsi Daerah Istimewa Yogyakarta, Jl. Brawijaya, Tamantirto, Kasihan, Bantul, 55183, Indonesia

\*Corresponding author: [20611141@students.uii.ac.id](mailto:20611141@students.uii.ac.id)



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

### Riwayat Artikel

Dikirim: 19 September 2024  
Direvisi: 16 Januari 2025  
Diterima: 27 Januari 2025

### ABSTRAK

Pertanian merupakan kegiatan yang sangat penting, tidak hanya sebagai penghasil pangan untuk kebutuhan pokok, tetapi juga memberikan kontribusi besar terhadap perekonomian. Penelitian ini bertujuan untuk mengoptimalkan potensi usaha perorangan pada subsektor pertanian di tiap kecamatan di Provinsi D.I Yogyakarta. Data yang digunakan dalam penelitian ini yaitu data jumlah usaha pertanian perorangan menurut kecamatan dan subsektor di provinsi D.I Yogyakarta 2023 yang diambil di publikasi Hasil Pencacahan Lengkap Sensus Pertanian D.I Yogyakarta. Hasil dari statiska deskriptif menunjukkan bahwa subsektor Perkebunan dan Jasa Pertanian memiliki jumlah usaha terendah, sementara subsektor Tanaman Pangan memiliki jumlah usaha tertinggi dengan nilai maksimal 10.481. Subsektor Hortikultura memiliki rata-rata tertinggi, sedangkan subsektor Kehutanan menunjukkan variasi jumlah usaha yang signifikan. Berdasarkan analisis *clustering* menggunakan *k-means*, terbentuk enam cluster dengan karakteristik unik. *Cluster 1* menonjol dengan aktivitas pertanian kuat dan beragam, sementara *Cluster 2* menghadapi tantangan signifikan. *Cluster 3* dan 4 memiliki nilai sedang dan seimbang, menunjukkan potensi baik dengan kebutuhan dukungan tambahan. *Cluster 5* unggul pada subsektor Tanaman Pangan, Peternakan, dan Kehutanan, sementara *Cluster 6* memiliki kekuatan di subsektor Hortikultura, Perkebunan, dan Kehutanan, menjadikannya cluster tertinggi. Disarankan pemerintah provinsi D.I Yogyakarta memfokuskan dukungan spesifik pada masing-masing cluster untuk memaksimalkan potensi pertumbuhan. *Cluster 1* memerlukan dukungan berkelanjutan, *Cluster 2* memerlukan evaluasi mendalam, sementara *Cluster 3, 4, 5, dan 6* membutuhkan program-program terarah sesuai sektor unggulan mereka.

**Kata Kunci:** Pertanian, Subsektor, Provinsi D.I Yogyakarta, Statistika Deskriptif, *K-means*.

## **ABSTRACT**

*Agriculture is a crucial activity, not only for producing food for basic needs but also for making a significant contribution to the economy. This study aims to optimize the potential of individual agricultural enterprises in each subdistrict of D.I Yogyakarta Province. The data used in this research includes the number of individual agricultural enterprises by subdistrict and subsector in D.I Yogyakarta Province for the year 2023, obtained from the publication "Hasil Pencacahan Lengkap Sensus Pertanian D.I Yogyakarta." Descriptive statistics results show that the Plantation and Agricultural Services subsectors have the lowest number of enterprises, while the Food Crops subsector has the highest number of enterprises with a maximum value of 10,481. The Horticulture subsector has the highest average, while the Forestry subsector shows significant variation in the number of enterprises. Based on clustering analysis using k-means, six clusters with unique characteristics were formed. Cluster 1 stands out with strong and diverse agricultural activities, while Cluster 2 faces significant challenges. Clusters 3 and 4 have moderate and balanced values, indicating good potential with a need for additional support. Cluster 5 excels in the Food Crops, Livestock, and Forestry subsectors, while Cluster 6 has strengths in the Horticulture, Plantation, and Forestry subsectors, making it the highest-ranking cluster. It is recommended that the D.I Yogyakarta provincial government focuses specific support on each cluster to maximize growth potential. Cluster 1 requires ongoing support, Cluster 2 needs in-depth evaluation, while Clusters 3, 4, 5, and 6 require targeted programs according to their leading sectors.*

*Keywords: Agriculture, Subsectors, D.I Yogyakarta Province, Descriptive Statistics, K-means.*

## **1. Pendahuluan**

Indonesia dikenal sebagai negara agraris dengan luas lahan yang sangat besar, yang menyediakan banyak peluang bagi masyarakat untuk berprofesi di sektor pertanian [1]. Selain berfungsi sebagai sumber mata pencaharian, sektor pertanian juga memiliki peran strategis dalam meningkatkan perekonomian negara. Komoditas pertanian Indonesia memiliki daya saing yang cukup tinggi di pasar internasional, menegaskan pentingnya sektor ini dalam konteks global.

Pertanian bukan hanya merupakan sektor ekonomi dasar, tetapi juga menjadi pondasi utama yang memengaruhi kestabilan ekonomi sebuah bangsa. Pembangunan sektor pertanian yang efektif berkontribusi langsung terhadap stabilitas ekonomi dan

pertumbuhan nasional [2]. Sebuah negara dapat dianggap maju jika mampu memenuhi kebutuhan pangan dasar rakyatnya secara memadai.

Di Provinsi Daerah Istimewa Yogyakarta, sektor pertanian memiliki potensi yang signifikan dengan adanya Lahan Pertanian Pangan Berkelanjutan seluas 35.911,59 hektar [3]. Oleh karena itu, penting untuk mengoptimalkan kebijakan yang mendukung sektor pertanian, khususnya dalam meningkatkan jumlah usaha pertanian perorangan. Pengembangan kebijakan yang tepat akan berkontribusi pada pertumbuhan sektor pertanian dan, pada gilirannya, akan memperkuat perekonomian daerah serta meningkatkan kesejahteraan masyarakat.

Sektor pertanian, beserta berbagai subsektornya, memainkan peranan krusial dalam struktur ekonomi Provinsi Daerah Istimewa Yogyakarta. Provinsi ini terbagi menjadi 78 kecamatan, masing-masing dengan karakteristik unik dalam usaha pertanian perorangan yang melibatkan berbagai subsektor. Kontribusi sektor pertanian terhadap ekonomi lokal sangat signifikan, sehingga pemahaman yang mendalam mengenai distribusi dan pengelompokan usaha pertanian perorangan menjadi kunci untuk merumuskan kebijakan yang efektif dan relevan.

Dengan tujuan tersebut, penulis akan melakukan analisis menggunakan metode *K-Means Clustering* terhadap kecamatan-kecamatan di Provinsi D.I Yogyakarta. Tujuan utama dari penerapan metode klasterisasi ini adalah untuk mengoptimalkan pengelompokan data dengan cara mengurangi variasi di dalam setiap kelompok [4]. Analisis ini bertujuan untuk mengoptimalkan pengelompokan usaha pertanian serta meningkatkan pemahaman tentang karakteristik masing-masing kelompok. Hasil pengelompokan ini akan membantu dalam pengambilan keputusan kebijakan dengan cara mengidentifikasi wilayah yang membutuhkan perhatian khusus, baik dalam hal peningkatan produktivitas, dukungan infrastruktur, atau program pelatihan bagi petani. Dengan cara ini, diharapkan dapat diperoleh wawasan yang lebih tajam mengenai pola distribusi dan pengelompokan usaha, yang akan mendukung perumusan kebijakan yang lebih terarah dan berbasis data.

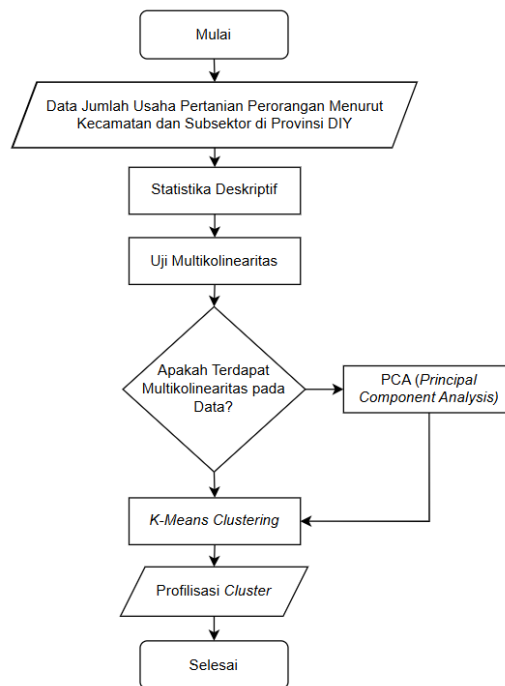
## **2. Metodologi Penelitian**

### **2.1. Data dan Sumber Data**

Data yang digunakan dalam analisis ini adalah data sekunder, yaitu data Jumlah Usaha Pertanian Perorangan menurut Kecamatan dan Subsektor di Provinsi D.I Yogyakarta (unit) pada tahun 2023 yang diperoleh dari Badan Pusat Statistik Provinsi D.I Yogyakarta.

### **2.2. Metodologi**

Analisis data dilakukan dengan menggunakan metode *k-means clustering* untuk mengoptimalkan jumlah usaha pertanian perorangan persubsektor seluruh kecamatan di provinsi D.I Yogyakarta tahun 2023. Analisis ini dibantu dengan menggunakan *software R Studio* dan *Microsoft Excel*. Proses analisis ini dilakukan dengan melewati beberapa tahapan untuk mencapai tujuan yang diinginkan. Berikut merupakan diagram alir penelitian yang akan menginformasikan tahapan-tahapan apa saja selama penelitian ini berlangsung:



**Gambar 2. 1.** Diagram Alir Penelitian

Berdasarkan diagram alir diatas, uraian tahapan yang dilakukan adalah sebagai berikut:

1. Mulai
2. Data yang digunakan dalam penelitian ini yaitu data jumlah usaha pertanian perorangan menurut kecamatan dan subsektor di provinsi D.I Yogyakarta 2023 yang diambil di publikasi Hasil Pencacahan Lengkap Sensus Pertanian D.I Yogyakarta
3. Melihat gambaran dan ringkasan data dengan menggunakan statistika deskriptif
4. Melakukan Uji *Multikolinearitas* untuk mengetahui apakah terdapat *multikolinearitas* atau korelasi yang tinggi antara dua atau lebih variabel independen.
5. Jika tidak terdapat *multikolinearitas*, maka analisis dapat dilanjutkan. Jika terdapat, maka harus mereduksi *multikolinearitas* dengan menggunakan *principal component analysis* (PCA)
6. Mengelompokkan kecamatan-kecamatan dengan menggunakan metode *k-means clustering*
7. Profilisasi untuk memberikan pemahaman yang lebih lanjut tentang karakteristik *cluster-cluster* yang terbentuk
8. Selesai

## 2.3. Landasan Teori

### 2.3.1. Pertanian

Pertanian adalah kegiatan yang dilakukan oleh manusia yang meliputi bercocok tanam, peternakan, perikanan, dan kehutanan [5]. Aktivitas ini melibatkan pemanfaatan bahan anorganis yang diolah dengan bantuan tumbuhan dan hewan yang memiliki kemampuan reproduksi, serta dilengkapi dengan upaya pelestarian. Pertanian dipandang

sebagai sebuah sistem ruang yang terdiri dari dua subsistem yaitu subsistem fisik dan subsistem manusia. Subsistem fisik mencakup unsur-unsur seperti tanah, iklim, hidrologi, dan topografi, bersama dengan proses-proses alamiah yang terjadi di dalamnya. Di sisi lain, subsistem manusia melibatkan tenaga kerja, kapasitas ekonomi, dan kondisi politik di daerah tersebut [6].

Pertanian adalah industri utama yang melibatkan pengaturan sumber daya seperti tanah, air, mineral, serta berbagai bentuk modal, dan pengelolaan tenaga kerja[7]. Keberhasilan proses ini dalam mencapai pengembangan pertanian sangat bergantung pada peran sumber daya manusia sebagai pelaksananya. Dalam sektor pertanian, sebagian besar usaha pertanian dikelola oleh pertanian rakyat.

### 2.3.2. Statistika Deskriptif

Statistika deskriptif adalah bidang dalam statistika yang berfungsi untuk mengumpulkan, menyusun, dan mengolah data sehingga dapat disajikan dan memberikan gambaran yang jelas tentang suatu kondisi atau peristiwa tertentu dari data yang diperoleh. Dengan kata lain, peran statistika deskriptif adalah untuk menampilkan data secara jelas agar dapat dipahami atau diinterpretasikan berdasarkan informasi yang disajikan [8].

### 2.3.3. Uji Multikolinearitas

*Multikolinearitas* adalah kondisi yang muncul ketika terdapat hubungan linear yang kuat atau sempurna antara beberapa atau bahkan semua variabel dalam model. *Multikolinearitas* terjadi saat lebih dari satu hubungan linear yang pasti terbentuk di antara variabel-variabel tersebut. Salah satu cara untuk mendeteksi *multikolinearitas* adalah dengan menghitung nilai *Variance Inflation Factor* (VIF) menggunakan rumus tertentu. Adapun rumus yang digunakan adalah sebagai berikut:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (1)$$

Rumus di atas adalah formula untuk menghitung VIF dari variabel independen ke- $j$  dalam analisis regresi. VIF mengukur tingkat multikolinearitas, yaitu korelasi antar variabel independen. Rumus tersebut menunjukkan bahwa VIF dari variabel  $j$  dapat dihitung sebagai kebalikan dari  $(1 - R_j^2)$ , dimana  $R_j^2$  adalah koefisien determinasi hasil regresi dari variabel independen  $j$  terhadap variabel independen lainnya. Nilai VIF yang tinggi menunjukkan adanya multikolinearitas yang kuat, yang dapat mempengaruhi stabilitas koefisien dalam model regresi. Nilai ambang VIF yaitu  $(VIF_j \geq 10)$ . Jika ditemukan indikasi *multikolinearitas*, maka perlu diambil tindakan untuk mengatasinya [9]

## 2.4. Principal Component Analysis (PCA)

*Principal Component Analysis* (PCA) merupakan teknik statistik multivariat yang mengubah sekelompok variabel asli secara linear menjadi sejumlah variabel baru yang lebih sedikit dan tidak berkorelasi, namun tetap mempertahankan informasi dari variabel awal [10]. Tujuan utama PCA adalah untuk menjelaskan sebanyak mungkin variansi dari data asli dengan menggunakan jumlah komponen utama yang minimal, yang disebut faktor. Menurut Ritonga dan Muhandhis, PCA adalah salah satu metode paling populer untuk pengurangan dimensi dan merupakan teknik statistik standar yang dapat digunakan untuk menyederhanakan dimensi suatu kumpulan data [11]. PCA menghasilkan komponen utama yang diperoleh dari dekomposisi *eigen value* dan *eigen vector* dari matriks kovarians. Berikut merupakan rumus-rumus yang diperlukan dalam PCA [12]:

Menghitung *mean* ( $\bar{X}$ ) dari tiap dimensi data dengan menggunakan persamaan:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

Rumus menghitung rata-rata dari suatu kumpulan data sampel adalah dengan menjumlahkan semua nilai data sampel ( $x_i$ ) dan kemudian membaginya dengan jumlah data sampel ( $n$ ).

Menghitung *covariance matrix* ( $C_x$ ) dengan menggunakan persamaan:

$$C_x = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \quad (3)$$

Rumus diatas merupakan rumus untuk menghitung matriks kovarians sampel ( $C_x$ ). Di sini,  $n$  adalah jumlah data sampel,  $x_i$  adalah vektor data sampel ke- $i$ , dan  $\bar{x}$  adalah vektor rata-rata sampel. Prosesnya melibatkan pengurangan vektor rata-rata  $\bar{x}$  dari setiap vektor data sampel  $x_i$ , mengalikan hasil selisih dengan transpose-nya sendiri, dan menjumlahkan semua hasil perkalian tersebut. Jumlah total ini kemudian dibagi dengan  $n - 1$  untuk mendapatkan matriks kovarians. Matriks kovarians ini mengukur seberapa banyak dua variabel berubah bersama-sama, yang berguna dalam analisis statistik dan machine learning.

Menghitung *eigenvector* ( $v_m$ ) dan *eigenvalue* ( $\lambda_m$ ) dari matriks kovarians dengan menggunakan persamaan:

$$C_x v_m = \lambda_m v_m \quad (4)$$

Persamaan diatas menunjukkan bahwa ketika matriks kovarians  $C_x$  diterapkan pada *eigenvector*  $v_m$ , hasilnya adalah *eigenvector*  $v_m$  yang diskalikan oleh *eigenvalue*  $\lambda_m$ . *Eigenvector* dan *eigenvalue* ini sangat penting dalam analisis *Principal Component Analysis* (PCA), untuk mengidentifikasi arah variabilitas maksimum dalam data.

### 2.5. K-Means Clustering

*Clustering* adalah teknik yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik di antara data [13]. Salah satu metode clustering *non-hierarki* adalah *K-Means*, yang bertujuan untuk membagi data ke dalam satu atau beberapa kelompok (*cluster*). Dalam metode ini, data dengan karakteristik yang mirip akan dikelompokkan ke dalam *cluster* yang sama, sedangkan data yang memiliki karakteristik berbeda akan ditempatkan pada *cluster* yang berbeda [14]. Berikut adalah langkah-langkah dalam menerapkan *k-means clustering* [15]:

1. Tentukan jumlah *cluster*  $k$  yang ingin dibentuk dan pilih posisi *centroid* awal secara acak dari objek yang ada.
2. Hitung jarak antara setiap data *input* dan *centroid* menggunakan rumus jarak *Euclidean*, kemudian identifikasi jarak terdekat antara setiap data dan *centroid*. Rumus Jarak *Euclidean* adalah:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

Di mana  $x_i$  dan  $y_i$  adalah koordinat dari dua titik dalam dimensi ke- $i$ . Rumus diatas mengukur jarak tengah antara dua titik

3. Hitung nilai *centroid* yang baru dengan mengambil rata-rata dari data dalam setiap *cluster*.
4. Hitung kembali jarak antara setiap objek dan *centroid* yang baru, mengikuti langkah 2.
5. Ulangi langkah 2 dan 3 sampai tidak ada perubahan dalam anggota *cluster*. Setelah proses ini selesai, nilai rata-rata pusat *cluster* ( $\mu_j$ ) pada iterasi terakhir akan digunakan untuk menentukan klasifikasi data.

### 3. Hasil dan Pembahasan

#### 3.1. Analisis Deskriptif

Analisis deskriptif digunakan pada pengumpulan, pengolahan, penyajian, dan pengambilan kesimpulan dari suatu kumpulan data. Tujuan utama dari statistika deskriptif adalah untuk memberikan gambaran yang jelas dan ringkas mengenai data yang dikumpulkan sehingga dapat dipahami dengan mudah. Analisis deskriptif diterapkan untuk melihat nilai minimal, mean dan maksimal. Hasil yang didapatkan dari analisis deskriptif adalah sebagai berikut.

**Tabel 3. 1.** Jumlah Usaha Perorangan Persubsektor Pertanian

Subsektor	Minimal	Mean	Maksimal
Tanaman Pangan	2	3.695	10.481
Hortikultura	7	2349	6.867
Perkebunan	0	1.467,4	7.976
Peternakan	25	3.952	9.562
Perikanan	7	301,86	1.148
Kehutanan	1	2.036,7	7.366
Jasa Pertanian	0	86,42	717

Berdasarkan **Tabel 3. 1**, menunjukkan mengenai jumlah usaha perorangan persubsektor pertanian di provinsi D.I Yogyakarta, terlihat bahwa nilai minimal pada subsektor Perkebunan dan Jasa Pertanian dengan nilai 0, yang menunjukkan bahwa ada kecamatan di mana tidak ada usaha di subsektor ini. Sedangkan nilai maksimal tertinggi berada pada subsektor Tanaman Pangan dengan nilai 10.481, yang menunjukkan jumlah usaha perorangan tertinggi. Nilai rata-rata tertinggi juga terdapat pada subsektor Hortikultura dengan 2.349, menunjukkan bahwa sebagian besar usaha berada di bidang ini. Rentang jumlah usaha pada subsektor Kehutanan adalah yang terluas, dari 1 hingga 7.366, menunjukkan variasi yang signifikan dalam jumlah usaha.

Secara keseluruhan, data ini menunjukkan bahwa subsektor Perkebunan dan Jasa Pertanian memiliki jumlah usaha terendah dan rentang yang paling kecil dibandingkan dengan subsektor lainnya, sementara subsektor Tanaman Pangan memiliki jumlah usaha dan rentang tertinggi. Subsektor Peternakan dan Kehutanan juga menunjukkan variasi yang cukup besar dalam jumlah usaha, namun tidak sebesar subsektor Tanaman Pangan. Subsektor Perikanan menunjukkan jumlah usaha yang cukup konsisten dengan nilai rata-rata yang mendekati subsektor Peternakan.

#### 3.2. Uji *Multikolinearitas*

Untuk menguji adanya *multikolinearitas*, kita akan menggunakan nilai *Variance Inflation Factor* (VIF). Proses ini melibatkan pengukuran variabel-variabel independen melalui statistik tertentu seperti VIF untuk mengidentifikasi dan mengatasi potensi masalah *multikolinearitas*. *Multikolinearitas* dikatakan terjadi apabila nilai VIF melebihi angka 10. Di bawah ini adalah hasil perhitungan nilai VIF antar variabel.

**Tabel 3. 2.** Nilai VIF

Tanaman Pangan	Hortikultura	Perkebunan	Peternakan	Perikanan	Kehutanan	Jasa Pertanian
14,204	6,480	5,025	19,569	1,799	6,195	1,237

Dari



**Tabel 3. 2** mengenai nilai *Variance Inflation Factor* (VIF) antar variabel, terlihat bahwa beberapa variabel menunjukkan nilai VIF di atas 10, yang menunjukkan adanya hubungan yang cukup kuat antara variabel-variabel tersebut. Khususnya, pada variabel Tanaman Pangan dan Peternakan yang masing-masing memiliki nilai VIF 14,204 dan 19,569. Berdasarkan hal tersebut, dapat ditarik kesimpulan bahwa terjadi *multikolinearitas*. Oleh karena itu, dilakukan *Principal Component Analysis* (PCA) untuk mengurangi *multikolinearitas*.

### 3.3. *Principal Component Analysis* (PCA)

*Principal Component Analysis* (PCA) adalah metode statistik multivariat yang mengubah serangkaian variabel asli menjadi serangkaian variabel yang lebih sedikit dan saling tidak berkorelasi, sambil tetap mempertahankan informasi dari variabel asli. Tujuan utamanya adalah untuk menjelaskan sebanyak mungkin variasi dari data asli menggunakan sejumlah komponen utama yang minimal, yang dikenal sebagai faktor. Untuk melakukan analisis PCA, perlu dilakukan uji kelayakan data dan kecukupan data dengan metode *Kaiser-Meyer-Olkin* (KMO) dan penentuan komponen utama dengan *eigen value*.

#### 3.3.1. *Kaiser-Meyer-Olkin* (KMO)

Berdasarkan uji *multikolinearitas* dapat diketahui bahwa data terdapat *multikolinearitas*. Oleh karena itu, dilakukan analisis PCA untuk mereduksi *multikolinearitas*. Sebelum melakukan analisis PCA, dibutuhkan uji kelayakan data dan kecukupan model. Uji ini dapat dilakukan dengan menggunakan metode *Kaiser-Meyer-Olkin* (KMO). Berikut merupakan tabel hasil uji KMO:

**Tabel 3. 3.** Uji KMO

Tanaman Pangan	Hortikultura	Perkebunan	Peternakan	Perikanan	Kehutanan	Jasa Pertanian	Overall
0,59	0,65	0,51	0,64	0,14	0,73	0,91	0,61

Dari **Tabel 3. 3** yang menunjukkan nilai MSA dari hasil uji KMO. Dapat diketahui bahwa *overall* MSA (*Measure of Sampling Adequacy*) dari model melebihi 0,5 sehingga data layak untuk diuji menggunakan PCA. Tetapi, hasil uji KMO variabel nya terdapat 1 variabel nilai MSA nya kurang dari 0,5 yaitu variabel perikanan. Nilai MSA digunakan untuk mengukur kecukupan sampel dalam analisis faktor. MSA memberikan indikator sejauh mana variabel-variabel dalam analisis faktor dapat diprediksi dan cocok untuk analisis faktor. Sehingga variabel perikanan di buang dari data karena akan mempengaruhi hasil dari PCA.

#### 3.3.2. Menentukan Komponen Utama dengan *Eigen Value*

Dengan menggunakan 6 Variabel tersisa, analisis PCA dapat dilanjutkan dengan langkah berikutnya yaitu penentuan *eigen value* untuk menentukan berapa komponen utama yang sudah cukup untuk menjelaskan keseluruhan data. Berikut merupakan tabel dari penentuan komponen utama menggunakan *eigen value*

**Tabel 3. 4.** Analisis PCA

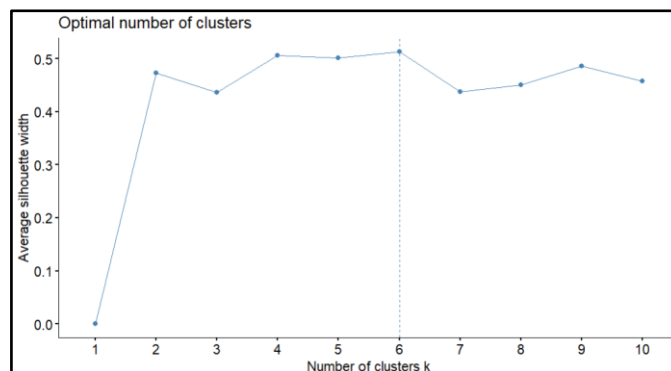
<i>Component</i>	<i>Eigen value</i>	<i>Percentage of Variance</i>	<i>Cumulative of Variance</i>
<i>Component 1</i>	3.815	63,58	63,58%
<i>Component 2</i>	1.110	18,50	82,09%
<i>Component 3</i>	0.663	11,05	93,14%
<i>Component 4</i>	0.215	03,58	96,73%
<i>Component 5</i>	0.167	02,78	99,52%
<i>Component 6</i>	0.028	00,47	100%

Hasil analisis PCA menunjukkan bahwa dua komponen utama sudah cukup untuk menjelaskan sebagian besar variansi dalam data. Komponen pertama menjelaskan 63,58%

variansi, sedangkan komponen kedua menambahkan 18,50%, sehingga totalnya menjadi 82,09%. Penggunaan dua komponen ini diusulkan karena mereka menangkap informasi yang signifikan sambil mempertahankan kesederhanaan dan efisiensi analisis. Hal ini penting untuk memastikan bahwa sebagian besar informasi dalam data tetap terwakili dengan baik, sambil mengurangi kompleksitas model. Pemilihan jumlah komponen didasarkan pada variansi yang dijelaskan oleh masing-masing komponen, di mana dua komponen utama tersebut telah mencakup lebih dari 80% variansi total, yang merupakan ambang batas yang umumnya diterima untuk analisis PCA yang efektif. Ini menunjukkan bahwa dua komponen ini cukup untuk memberikan gambaran yang akurat tentang struktur data, tanpa perlu menambahkan komponen tambahan yang hanya memberikan sedikit kontribusi terhadap variansi total.

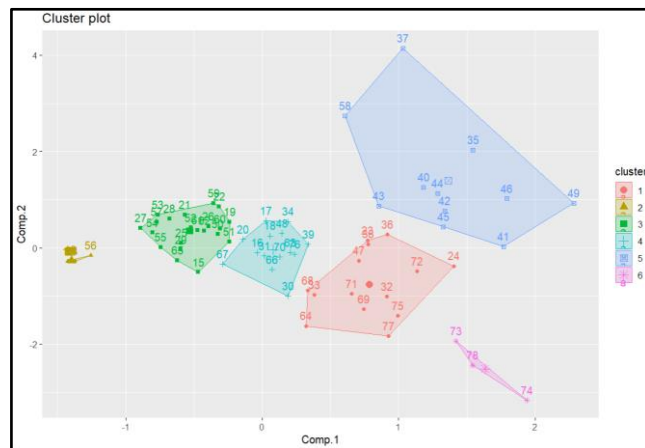
### 3.4. K-Means

Dengan menggunakan data yang telah dianalisis melalui PCA, langkah selanjutnya adalah mengoptimalkan jumlah *cluster*. Pada algoritma *K-Means*, sangat penting untuk memastikan bahwa data dikelompokkan dengan cara yang paling bermakna dan representatif. *K-Means* dipilih karena keunggulannya dalam efisiensi komputasi, data dengan distribusi yang jelas, dan kemampuannya memberikan hasil yang mudah diinterpretasikan. Pemilihan jumlah *cluster* yang tepat memungkinkan kita untuk menangkap struktur alami dalam data, meningkatkan akurasi analisis, dan menghasilkan wawasan yang lebih mendalam. Oleh karena itu, metode *Silhouette* digunakan untuk menentukan jumlah *cluster* yang optimal. Berikut *plot* yang didapat:



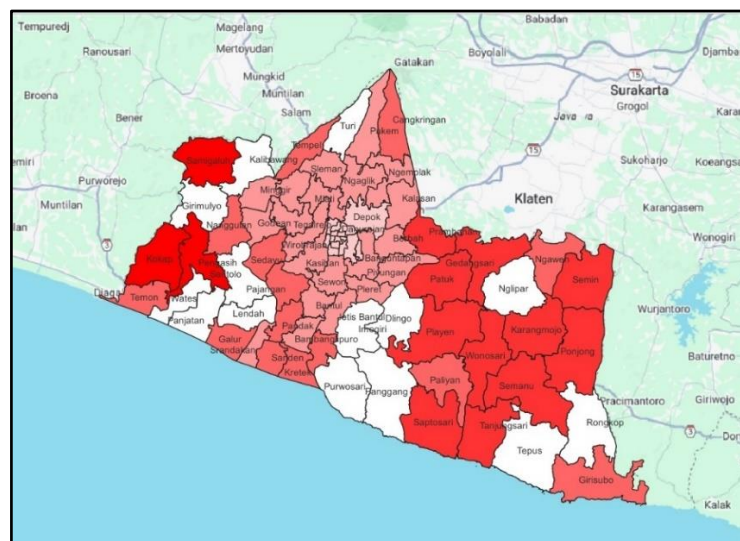
Gambar 3. 1. Silhouette Optimization Plot

Pada Gambar *Silhouette Optimization Plot*, dapat dilihat bahwa garis putus-putus yang terbentuk pada metode *Silhouette* menunjukkan bahwa jumlah cluster optimal adalah 6, dengan nilai koefisien *Silhouette* sebesar sekitar 0.5. Ini berarti bahwa enam *cluster* adalah jumlah yang optimal untuk mengelompokkan data secara bermakna dan representatif. Nilai koefisien *Silhouette* yang tinggi menunjukkan bahwa objek dalam *cluster* sangat mirip satu sama lain dan berbeda dengan objek di cluster lain. Pemilihan jumlah *cluster* ini penting untuk memastikan bahwa struktur alami dalam data ditangkap dengan baik, sehingga analisis selanjutnya dapat dilakukan dengan lebih efektif dan memberikan wawasan yang lebih tajam. Dengan menggunakan jumlah 6 *cluster*, maka akan dilakukan pembentukan 6 *cluster* dengan menggunakan metode *K-Means* dan ditampilkan visualisasi sebaran tiap *cluster* nya. Berikut visualisasinya.



Gambar 3. 2. Visualisasi Pengelompokan Cluster dengan K-Means

Pada Gambar 3. 2, terlihat bahwa terdapat 6 cluster pengelompokan kecamatan berdasarkan analisis yang dilakukan. Cluster-cluster ini memiliki distribusi yang berbeda-beda. Cluster 1 terdiri dari 14 kecamatan. Cluster 2 mencakup 15 kecamatan. Cluster 3 merupakan cluster dengan jumlah kecamatan paling banyak, yaitu sebanyak 21 kecamatan. Cluster 4 juga terdiri dari 14 kecamatan. Cluster 5 mencakup 11 kecamatan, sedangkan Cluster 6 terdiri dari 3 kecamatan dan memiliki jumlah kelurahan paling sedikit. Untuk penjelasan lebih rinci mengenai kecamatan mana saja yang termasuk dalam setiap cluster, dapat dilihat pada gambar dan tabel berikut:



Gambar 3. 3. Visualisasi Cluster pada peta

Tabel 3. 5. Keterangan Kecamatan Tiap Cluster

Cluster	Jumlah Anggota	Kecamatan
1	14	Imogiri, Dlingo, Panggang, Purwosari, Tepus, Rongkop, Nglipar, Turi, Wates, Panjatan, Lendah, Sentolo, Girimulyo, Kalibawang
2	15	Mantrijeron, Kraton, Mergangsan, Umbulharjo, Kotagede, Gondokusuman, Danurejan, Pakualaman, Gondomanan, Ngampilan, Wirobrajan, Gedongtengen, Jetis Kota, Tegalrejo, Depok
3	21	Srandakan, Bambanglipuro, Bantul, Jetis Bantul, Pleret, Piyungan, Banguntapan, Sewon, Kasihan, Moyudan, Minggir, Seyegan, Godean, Gamping, Mlati, Berbah, Kalasan, Ngemplak, Ngaglik, Sleman, Pakem

Cluster	Jumlah Anggota	Kecamatan
4	14	Sanden, Kretek, Pundong, Pandak, Pajangan, Sedayu, Paliyan, Girisubo, Ngawen, Tempel, Cangkringan, Temon, Galur, Nanggulan
5	11	Saptosari, Tanjungsari, Semanu, Ponjong, Karangmojo, Wonosari, Playen, Patuk, Gedangsari, Semin, Prambanan
6	3	Pengasih, Kokap, Samigaluh

### 3.5. Profilisasi

Setelah proses pengelompokan data (*clustering*) menggunakan metode *k-means* selesai dilakukan, langkah berikutnya adalah melakukan *profiling* atau profilisasi pada setiap *cluster* yang terbentuk. Profilisasi *cluster* ini bertujuan untuk mengidentifikasi dan memahami karakteristik unik dari masing-masing *cluster*. Melalui proses ini, kita dapat menggali dan mengungkap pola, kecenderungan, serta atribut-atribut dominan yang ada dalam setiap kelompok data. Dengan demikian, profilisasi *cluster* tidak hanya membantu dalam memahami perbedaan antar *Cluster*, tetapi juga memberikan wawasan yang lebih dalam mengenai data yang dianalisis.

**Tabel 3. 6.** Profilisasi

Cluster	Tanaman Pangan	Hortikultura	Perkebunan	Peternakan	Kehutanan	Jasa Pertanian	Mean
1	4690,5	4088,5	3019,5	5617,929	3852,857	56,785	3554.345
2	64,2	59,8	6,8	181,2	5,733	1,4	53.188
3	3081,095	1540	471,285	2858,667	428,381	87,523	1411.159
4	4054,357	2650,071	1744,5	4319,786	1522,571	89,928	2396.869
5	8145,909	3436,273	1539	7906,909	5112,727	245,545	4397.727
6	3513,667	5939,333	6944,667	6455,333	6094,333	42,333	4831.611

Tabel **Tabel 3. 6** menyajikan profilisasi *cluster* dari usaha perorangan berdasarkan subsektor pertanian di Provinsi D.I. Yogyakarta. Ada enam *cluster* yang terbentuk berdasarkan nilai dari enam subsektor, yaitu Tanaman Pangan, Hortikultura, Perkebunan, Peternakan, Kehutanan, dan Jasa Pertanian. *Cluster 1* menonjol dengan nilai yang tinggi pada hampir semua subsektor, menunjukkan bahwa daerah-daerah dalam *cluster* ini memiliki aktivitas pertanian yang kuat dan beragam dan berada dengan urutan ke-3 tertinggi. Sebaliknya, *cluster 2* menunjukkan nilai yang terendah untuk hampir semua subsektor, menandakan bahwa daerah-daerah dalam *cluster* ini mungkin menghadapi tantangan signifikan dalam pengembangan pertanian.

*Cluster 3* dan *cluster 4* memiliki nilai yang sedang dan menunjukkan variasi yang lebih seimbang di antara subsektor, menunjukkan bahwa daerah-daerah ini memiliki potensi yang cukup baik tetapi mungkin memerlukan dukungan tambahan untuk mengoptimalkan produksi. *Cluster 5* menunjukkan nilai yang sangat tinggi pada beberapa subsektor utama seperti Tanaman Pangan, Peternakan, dan Kehutanan, sehingga *cluster* ini mungkin memiliki spesialisasi yang kuat dalam bidang-bidang tersebut dan menjadi *cluster* urutan ke-2 tertinggi. *Cluster 6* memiliki kekuatan khusus di subsektor Hortikultura, Perkebunan, dan Kehutanan, menandakan potensi besar dalam ketiga subsektor tersebut. Tetapi, *cluster 6* memiliki nilai rata-rata tertinggi di setiap subsektornya dan menjadikan *cluster 6* sebagai *cluster* tertinggi.

Berdasarkan analisis ini, pemerintah provinsi D.I. Yogyakarta disarankan untuk memfokuskan dukungan dan program pengembangan yang spesifik pada masing-masing *cluster*. *Cluster 1* memerlukan dukungan berkelanjutan untuk mempertahankan dan meningkatkan produksi, sementara *Cluster 2* membutuhkan evaluasi mendalam untuk mengidentifikasi kendala dan peluang peningkatan. *Cluster 3* dan *Cluster 4* dapat dioptimalkan dengan program-program yang terarah sesuai dengan kebutuhan mereka.

*Cluster 5* dan *Cluster 6* memerlukan perhatian khusus pada subsektor-sektor dimana mereka menunjukkan kekuatan, untuk memaksimalkan potensi pertumbuhan yang ada.

#### 4. Kesimpulan

Dari hasil penelitian yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

1. Berdasarkan analisis *cluster* terhadap jumlah usaha perorangan di subsektor pertanian Provinsi D.I Yogyakarta 2023, beberapa tindakan strategis dapat dilakukan. Untuk subsektor Perkebunan dan Jasa Pertanian yang memiliki jumlah usaha rendah, perlu adanya dukungan berupa bantuan teknis dan program insentif untuk mendorong pertumbuhan usaha di sektor ini, terutama di kecamatan yang belum memiliki usaha serupa. Subsektor Tanaman Pangan yang mendominasi usaha mencapai 10.481. Oleh karena itu, perlu diperkuat dengan fasilitas penyimpanan dan pengolahan hasil panen untuk meningkatkan nilai tambah. Di subsektor Hortikultura, yang memiliki nilai rata-rata tinggi, fokus pada pengembangan kualitas produk dan pemasaran premium dapat meningkatkan daya saing di pasar lokal dan nasional. Untuk subsektor Kehutanan, yang memiliki variasi signifikan dengan rentang dari 1 hingga 7.366, upaya pemberdayaan masyarakat melalui program hutan sosial dapat menambah pendapatan dan memastikan pengelolaan yang berkelanjutan. Terakhir, subsektor Perikanan dan Peternakan yang cukup konsisten dapat dioptimalkan melalui pelatihan berkelanjutan dan penguatan akses pasar. Pendekatan ini diharapkan dapat memperkuat setiap subsektor sesuai dengan karakteristik dan potensi yang ada.
2. Berdasarkan hasil pengelompokan subsektor pertanian di Provinsi D.I Yogyakarta, berikut adalah rekomendasi kebijakan untuk setiap *cluster*. Untuk *Cluster 1* yang memiliki aktivitas pertanian kuat dan beragam, pemerintah dapat meningkatkan diversifikasi produk dan akses pasar. *Cluster 2*, yang menghadapi tantangan besar, membutuhkan bantuan khusus seperti subsidi pertanian dan penyuluhan untuk meningkatkan produktivitas. *Cluster 3* dan 4, dengan potensi yang seimbang, perlu dukungan teknologi dan modal usaha untuk optimalisasi. *Cluster 5*, dengan nilai tinggi di subsektor Tanaman Pangan, Peternakan, dan Kehutanan, dapat didukung dengan infrastruktur pasca-panen dan branding produk unggulan. Sementara itu, *Cluster 6* yang unggul di subsektor Hortikultura, Perkebunan, dan Kehutanan perlu didorong untuk pengembangan produk premium dan ekspor melalui bantuan sertifikasi dan promosi. Kebijakan yang fokus dan terarah ini diharapkan mampu mendorong perkembangan sektor pertanian di setiap *cluster* sesuai potensinya.

#### 5. Daftar Pustaka

- [1] H. M. Qudrotulloh, E. Sumarsih, H. Nuryaman, N. R. Mutiarasari, and T. Hardiyanto, "Persepsi Petani Muda Terhadap Wirausaha di Sektor Pertanian (Studi Kasus Pada Petani Muda di Desa Tenjonagara)," *Agribisnis dan Teknologi Pangan*, vol. 2, no. 2, pp. 124–135, 2022, doi: 10.32627.
- [2] R. D. Puspitasari, "Pertanian Berkelanjutan Berbasis Revolusi Industri 4.0," *Jurnal Layanan Masyarakat Universitas Airlangga*, vol. 03, no. 01, pp. 26–28, 2019, [Online]. Available: [https://id.wikipedia.org/wiki/Sukorame,\\_Lamongan](https://id.wikipedia.org/wiki/Sukorame,_Lamongan)
- [3] G. I. Janti, E. Martono, and Subejo, "Perlindungan Lahan Pertanian Pangan Berkelanjutan Guna Memperkokoh Ketahanan Pangan Wilayah (Studi di Kabupaten Bantul, Daerah Istimewa Yogyakarta)," *Jurnal Ketahanan Nasional*, vol. 22, pp. 1–21, 2016.

- [4] H. Priyatman, F. Sajid, and D. Haldivany, “Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa,” *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 2019.
- [5] L. N. Nasution and M. Yusuf, “Analisis Konsumsi, Ekspor, dan Pertanian Terhadap Pertumbuhan Ekonomi di Sumatera Utara,” *Jurnal Kajian Ekonomi dan Kebijakan Publik*, vol. 3, no. 2, 2018.
- [6] B. S. Primada and I. Zaki, “Tinjauan Mekanisme Kontrak Pengelolaan Lahan Pertanian Berbasis Adat Istiadat Dalam Kajian Fiqh Muamalah (Desa Temu, Kecamatan Kanor, Kabupaten Bojonegoro),” *Jurnal Ekonomi Syariah Teori dan Terapan*, vol. 2, no. 11, 2015.
- [7] M. Batubara and M. Mustika Pane, “Pengaruh Pertanian terhadap Pendapatan Nasional,” *Jurnal Penelitian Ekonomi Akuntansi (JENSI)*, vol. 7, no. 1, pp. 74–81, 2023.
- [8] L. D. Martias, “Statistika Deskriptif Sebagai Kumpulan Informasi,” *Fihris: Jurnal Ilmu Perpustakaan dan Informasi*, vol. 16, no. 1, p. 40, 2021, doi: 10.14421/fhrs.2021.161.40-59.
- [9] D. R. Ningrat, D. Asih, I. Maruddani, and T. Wuryandari, “Analisis Cluster Dengan Algoritma K-Means Dan Fuzzy C-Means Clustering Untuk Pengelompokan Data Obligasi Korporasi,” *Jurnal Gaussian*, vol. 5, no. 4, pp. 641–650, 2016.
- [10] M. Wangge, “Penerapan Metode Principal Component Analysis (PCA) Terhadap Faktor-faktor yang Mempengaruhi Lamanya Penyelesaian Skripsi Mahasiswa Program Studi Pendidikan Matematika FKIP UNDANA,” *Jurnal Cendekia : Jurnal Pendidikan Matematika*, vol. 5, no. 2, pp. 974–988, 2021, doi: 10.31004/cendekia.v5i2.465.
- [11] A. S. Ritonga and I. Muhandhis, “Teknik Data Mining Untuk Mengklasifikasikan Data Ulasan Destinasi Wisata Menggunakan Reduksi Data Principal Component Analysis (Pca),” *Eduatic - Scientific Journal of Informatics Education*, vol. 7, no. 2, 2021, doi: 10.21107/edutic.v7i2.9247.
- [12] D. Hedyati and I. M. Suartana, “Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro,” *Journal of Information Engineering and Educational Technology*, vol. 5, no. 2, pp. 49–54, 2021, doi: 10.26740/jieet.v5n2.p49-54.
- [13] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, “Analisis K-Means Clustering pada Data Sepeda Motor,” *INFORMAL: Informatics Journal*, vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.
- [14] B. Melpa Metisen and H. Latipa Sari, “Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila,” *Jurnal Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.
- [15] J. Hutagalung and F. Sonata, “Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah,” *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, p. 1187, 2021, doi: 10.30865/mib.v5i3.3113.