

OPTIMALISASI *CLUSTER* PADA *CLUSTER* HIERARKI MENGGUNAKAN *PSEUDO F-STATISTIC CALINSKI HARABASZ* UNTUK KETAHANAN PANGAN

Mita Nourma Maulina^{1*}, Mustafid², Bagus Arya Saputra³

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*e-mail : mitamaulina66@gmail.com

DOI: 10.14710/j.gauss.14.1.01-12

Article Info:

Received: 2024-06-21

Accepted: 2025-02-04

Available Online: 2025-02-12

Keywords:

Agglomerative Hierarchical Clustering; GUI Python, Food Security System; Calinski Harabasz Pseudo F-Statistic

Abstract: Food security is the ability of a nation to ensure that all of its inhabitants obtain food in sufficient quantities, of proper quality, based on optimizing the utilization of food. The food security of each region is different, resulting in a region to have low or high food security, one of which is the area in Pati Regency as a research area. To find out this, area clustering was carried out based on variables of planted area, food production, number of food supply facilities and infrastructure, number of people with low welfare levels, number of sub-districts without sufficient connecting access, number of households without access to clean water, area, number of residents, number of households, and level of settlement. The grouping of food security areas was carried out using the Hierarchical Clustering method based on Euclidean distance calculations with cluster validation using Calinski Harabasz Pseudo F-Statistics. The results of this study obtained the optimum number of clusters as 7 clusters in the Complete Linkage closest distance calculation technique which has the highest Pseudo-F value. High quality food security is found in the highlands, namely, Gembong District, Cluwak District, Trangkil District, and Gunung Wungkal District. Meanwhile, the low quality of food security is found in lowland areas, namely, Juwana District and Pati District.

1. PENDAHULUAN

Di tengah era globalisasi ini, negara-negara di seluruh dunia dihadapkan pada tantangan dan ancaman ketahanan pangan yang kompleks dan variatif. Ketahanan pangan menjadi isu multidimensi yang sangat kompleks (Suryana, 2014). Terdapat 3 komponen utama dalam ketahanan pangan, yaitu *food availability and stability*, *food accessibility*, dan pemanfaatan pangan. Kabupaten Pati memiliki 21 kecamatan dengan kondisi yang berbeda-beda dalam hal ketahanan pangan. Untuk mengetahui tingkat ketahanan pangan di kecamatan-kecamatan tersebut, dilakukan *clustering* menggunakan metode *Agglomerative Hierarchical Clustering*. Metode ini membangun hierarki dari simpul-simpul menjadi *cluster* yang lebih besar. Untuk menentukan metode penentuan jumlah *cluster* optimal, digunakan metode *Calinski Harabasz Pseudo F-Statistic*.

Pada penelitian Sudirman *et al.* (2018) tentang analisis ketahanan pangan terbatas pada metode *K-Means* dan memiliki keterbatasan dalam jumlah indikator pengukuran. Selain itu, penelitian oleh Kingrani *et al.* (2018) *Calinski Harabasz Pseudo F-Statistic* menjadi kriteria penetapan jumlah *cluster* optimum yang relatif baik untuk mengatasi data *outlier*. Berdasarkan penelitian sebelumnya, penelitian ini akan membahas tentang *clustering* wilayah ketahanan pangan di Kabupaten Pati menggunakan metode *Agglomerative Hierarchical Clustering* dan optimalisasi *cluster* dengan *Calinski Harabasz Pseudo-F Statistic* menggunakan *GUI Python* untuk mempermudah perhitungan komputasi.

2. TINJAUAN PUSTAKA

Ketahanan pangan dapat dilihat dari adanya cukup pangan yang tersedia bagi rumah tangga, baik secara jumlah maupun kualitasnya. Konsep ini meliputi 4 aspek, yaitu *sufficiency, access, security, and time* (Baliwaty *et al.*, 2004). Untuk menganalisis tingkat ketahanan pangan suatu daerah, dapat dilakukan pengelompokan berdasarkan tiga komponen utama, yaitu ketersediaan dan *food stability*, aksesibilitas pangan, dan pemanfaatan pangan

Analisis *cluster* merupakan metode statistika amultivariat yang diterapkan untuk mengelompokkan data yang memiliki kesamaan ke dalam satu *cluster*. Data pada analisis *cluster* direpresentasikan sebagai *matrix* yang memiliki t kolom dan n , dimana setiap atribut memiliki satu kolom. Analisis *cluster* biasanya direpresentasikan sebagai a vektor pengukuran atau titik dalam ruang multidimensi (Kaur *et al.*, 2014).

Untuk mengatasi perbedaan skala data dalam analisis *cluster*, dilakukan proses standardisasi data. Standardisasi merupakan teknik yang mengubah data agar memiliki rata-rata 0 (terpusat) dan standar deviasi 1. Salah satu metode standardisasi yang umum digunakan adalah *Z-score* pada persamaan 1.

$$z = \frac{X - \mu}{\sigma} \quad (1)$$

dengan X adalah nilai data asli, μ yaitu nilai rata-rata dari data, dan σ merupakan nilai standar deviasinya.

Sebelum melakukan *clustering* perlu dilakukan pengujian asumsi, diantaranya sebagai berikut:

a. Asumsi Representativeness of The Sample

Asumsi *Representativeness of The Sample* bisa diperoleh dari uji *kaiser Mayer Olkin* (KMO) untuk menentukan sejauh mana sampel mewakili populasi. Uji KMO dilakukan untuk melihat validitas faktor-faktor dalam penelitian. Nilai KMO dapat diperoleh melalui Persamaan 2.

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (2)$$

dengan a_{ij} diperoleh dari rumus $a_{ij} = \frac{-r_{ij}}{\sqrt{r_{ij}r_{ij}}}$. Untuk p merupakan banyaknya variabel, r_{ij} yaitu tingkat keterkaitan antara variabel i dan j , sedangkan a_{ij} merupakan tingkat keterkaitan parsial antara variabel i dan j . Apabila nilai KMO $> 0,5$ maka dapat disimpulkan bahwa sampel yang diwakili memenuhi asumsi, sehingga memungkinkan untuk melanjutkan uji asumsi tahap selanjutnya (Andini, 2017).

b. Asumsi Non Multikolinieritas

Multikolinieritas mengacu pada adanya korelasi antara variabel dalam suatu objek. Untuk melakukan *clustering*, harus menghindari adanya korelasi antarobjek, jika terdapat korelasi antarobjek, disarankan untuk menghapus beberapa variabel yang tingkat bungannya cukup tinggi (Santoso, 2018). Untuk mengetahui ada atau tidaknya multikolinieritas dapat dilihat dari nilai VIF (*Variance Inflation Factor*). Jika nilai VIF < 10 maka tidak ada multikolinieritas.

$$VIF = \frac{1}{1 - R_{yx_1x_2}^2} \quad (3)$$

dengan $R_{yx_1x_2} = \sqrt{\frac{(r_{yx_1} - r_{yx_2})^2}{1 - r_{x_1x_2}^2}}$. Untuk $R_{yx_1x_2}^2$ merupakan koefisien determinasi. $R_{yx_1x_2}$ yaitu korelasi ganda antarvariabel y , x_1 , dan x_2 , serta r_{yx} merupakan korelasi antarvariabel x dan y . r_{yx} diperoleh dari Persamaan 4

$$r_{yx} = \frac{(n \sum y_i x_i - \sum y_i \sum x_i)}{\sqrt{(n \sum y_i - (\sum y_i)^2)(n \sum x_i^2 - (\sum x_i)^2)}} \quad (4)$$

Untuk n merupakan jumlah dari data yang digunakan.

Dalam mengatasi asumsi terjadinya multikolinieritas, dalam penelitian Johnson dan Wichern (2007), PCA (*Principal Component Analysis*) mampu menghasilkan variabel baru yang sudah terhindar dari masalah multikolinieritas. PCA akan mereduksi seluruh variabel yang digunakan menjadi komponen utama yang dapat mewakili seluruh variabel. Variabel baru yang terbentuk berdasarkan hasil PCA merupakan komponen utama yang dibentuk dari kombinasi linier variabel asalnya. Berikut disajikan prosedur dalam melakukan PCA:

a. Membentuk matriks data, misalkan matriks X dengan ordo $n \times m$ yang memperlihatkan nilai tertentu dari variabel ke- j terhadap objek ke- i , untuk $i = 1, 2, 3, \dots, n$ dan $j = 1, 2, 3, \dots, m$

b. Memilih matriks inputan

1. Matriks varian-kovarian (jika variabel dalam proses PCA memiliki ukuran yang sama)

$$S = \begin{bmatrix} s_{11}^2 & s_{12} & \dots & s_{1m} \\ s_{21} & s_{22}^2 & \dots & s_{2m} \\ \vdots & \vdots & \dots & \vdots \\ s_{m1} & s_{m2} & \dots & s_{m2}^2 \end{bmatrix} \quad (5)$$

$$s_j^2 = s_{jj} \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2 \quad (6)$$

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k) \quad (7)$$

dengan s_j^2 yaitu varian sampel untuk variabel ke- j sejumlah m variabel. Serta s_{jk} yaitu kovarian sampel untuk variabel ke- j dan ke- k pada sejumlah m variabel.

2. Matriks korelasi (apabila variabel dalam proses PCA memiliki satuan yang berbeda sehingga harus distarndardisasi terlebih dahulu)

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \vdots & \vdots & \dots & \vdots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix} \quad (8)$$

$$r_{jk} = \frac{\sum_{i=1}^n (z_{ij} - \bar{z}_j)(z_{ik} - \bar{z}_k)}{\sqrt{\sum_{i=1}^n (z_{ij} - \bar{z}_j)^2} \sqrt{\sum_{i=1}^n (z_{ik} - \bar{z}_k)^2}} \quad (9)$$

dengan r_{jk} merupakan sampel untuk variabel ke- j dan ke- k pada sejumlah m variabel.

c. Menentukan *eigen value* ($\lambda_1, \lambda_2, \dots, \lambda_p$) dari matriks inputan dapat dicari menggunakan persamaan karakteristik pada matriks inputan A (Anton dan Rorres, 2004).

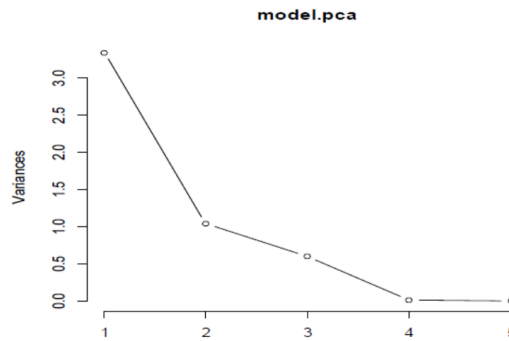
$$\det(\lambda I - A) = 0 \quad (10)$$

Setelah mendapatkan nilai *eigen*, kemudian menetapkan *eigen vector* ke- j untuk *eigen value* ke- j , dengan $j = 1, 2, \dots, p$. *Eigen vector* matriks A yang terikat pada *eigen value* λ adalah vektor-vektor tak nol yang memenuhi persamaan 11.

$$Ax = \lambda x \quad (11)$$

Persamaan 11 juga dapat ditulis menjadi $(\lambda I - A)x = 0$ yang merupakan ruang solusi dengan λ merupakan *eigen value* dari A serta untuk x adalah *vector eigen* dari A yang terikat dengan λ (Anton dan Rorres, 2004).

d. Menentukan jumlah komponen utama yang harus diambil, ada beberapa cara untuk mengidentifikasi jumlah komponen utama yang harus diambil diantaranya yaitu berdasarkan *scree plot*, jumlah komponen utama yang diambil yaitu pada titik kurva yang tidak lagi menurun tajam atau terbentuk siku (*elbow*) (Johnson dan Wichern, 2007).



Gambar 1. Contoh *Scree Plot*

Selain berdasarkan *scree plot*, banyak nya komponen utama bisa dilihat berdasarkan pada total varian data yang dapat dijelaskan secara kumulatif diantara 70% sampai 90% (Nugroho, 2008). Atau pemilihan komponen utama sesuai *eigen value* yang nilainya $\lambda_i > 1$ (Nugroho, 2008).

e. Menentukan model untuk komponen utamanya

1. Persamaan untuk matriks varian-kovarian

$$\begin{aligned} V_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1q}X_p \\ V_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2q}X_p \\ &\vdots \\ V_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pq}X_p \end{aligned}$$

(12)

2. Persamaan untuk matriks korelasi

$$\begin{aligned} C_1 &= a_{11}Z_1 + a_{12}Z_2 + \dots + a_{1q}Z_p \\ C_2 &= a_{21}Z_1 + a_{22}Z_2 + \dots + a_{2q}Z_p \\ &\vdots \\ C_p &= a_{p1}Z_1 + a_{p2}Z_2 + \dots + a_{pq}Z_p \end{aligned}$$

(13)

f. Menghitung skor komponen utama dengan mensubstitusikan nilai dari masing-masing data ke dalam model persamaan pada langkah ke-e. Skor komponen utama tersebut yang menjadi pengganti dari data variabel awal yang sudah terhindar dari masalah multikolinieritas.

Setelah data telah memenuhi semua asumsi, maka dapat dilakukan *clustering* data. *Clustering* adalah proses pengelompokan objek, catatan, atau kelas yang memiliki kesamaan atau kemiripan (Larose, 2015). Salah satu pendekatan dalam *clustering* adalah pendekatan hierarki, dimana data dikelompokkan berdasarkan kemiripan terdekat hingga membentuk pohon *cluster* (dendogram). Dalam pendekatan ini, perhitungan jarak antardata menggunakan metode *Euclidean*. Jarak observasi ke-*i* dan ke-*j* dapat dihitung melalui Persamaan 14.

$$d_{ik} = \sqrt{\sum_{j=1}^m (X_{ij} - X_{kj})^2}$$

(14)

dengan d_{ik} yaitu jarak objek ke-*i* dan objek ke-*k*, X_{ij} merupakan objek ke-*i* pada variabel ke-*j* dan X_{kj} yaitu nilai tengah sampel dalam kelompok ke-*j* dan variabel ke-*k*.

Algoritma *Hierarchical Clustering* bersifat *agglomerative (bottom-up)* dan *divisive (top-down)* (Han *et al.*, 2012). Dalam metode *agglomerative*, prosedurnya menghasilkan serangkaian partisi data, dimulai dari partisi yang terdiri dari objek tunggal hingga partisi yang terdiri dari satu kelompok yang mencakup semua objek. Metode ini memiliki beberapa teknik dalam perhitungan jarak terdekat, yaitu sebagai berikut (Andini, 2017).

a. *Single Linkage*

Jarak antarcluster didefinisikan sebagai jarak antara pasangan objek terdekat.

$$d_{(ij)k} = \min\{d_{ik}, d_{jk}\} \quad (15)$$

besaran d_{ik} dan d_{jk} merupakan jarak terpendek antara cluster – cluster i dan k , serta cluster j dan k (Johnson dan Winchern, 2007).

b. *Complete Linkage*

Penggabungan cluster yang didasarkan pada jarak maksimum antara dua cluster yang akan digabungkan.

$$d_{(ij)k} = \max\{d_{ik}, d_{jk}\} \quad (16)$$

dengan d_{ik} dan d_{jk} berturut – turut adalah jarak terpendek antara cluster – cluster i dan k , dan antara cluster – cluster j dan k (Johnson dan Winchern, 2007).

c. *Average Linkage*

Metode ini didasarkan pada jarak rata-rata antarobservasi. Pengelompokan dimulai dari pasangan observasi dengan jarak paling mendekati jarak rata-rata.

$$d_{(ij)k} = \frac{\sum_a \sum_b d_{ab}}{N_{ij}N_k} \quad (17)$$

dengan d_{ab} merupakan jarak antara objek a pada cluster ij dan objek b pada cluster k . Untuk N_{ij} yaitu banyaknya item pada cluster ij serta N_k yaitu banyaknya item pada cluster ij dan k (Johnson dan Winchern, 2007).

d. *Ward's Method*

Metode ini menghitung jarak antara dua kelompok dengan cara menjumlahkan kuadratkan error antara dua kelompok untuk semua variabel, untuk mendapatkna cluster yang memiliki varian dalam cluster sekecil mungkin (Andini, 2017).

$$SSE = \sum_{j=1}^N (x_j - \bar{x})^T (x_j - \bar{x}) \quad (18)$$

dengan x_j merupakan data dari objek ke- j dan \bar{x} yaitu rata-rata dari keseluruhan objek dalam cluster (Johndon & Wichern, 2007).

Metode *Agglomerative Hierarchical Clustering* mempunyai beberapa metode dalam menentukan jarak terdekat. Untuk menentukan metode penentuan jarak dan banyaknya cluster optimum, digunakan metode *Calinski Harabasz Pseudo F-Statistic*. Miligan dan Cooper melakukan penelitian pada tahun 1985 membuktikan bahwa *Pseudo F-statistic* memperoleh kinerja terunggul diantara 30 tekni yang tersedia, serta mampu digunakan secara global. *Pseudo F* terbesar membuktikan bahwa kelompok tersebut dapat memberikan hasil terbaik, dengan keragaman pada kelompok bersifat homogen sedangkan antarkelompok yang heterogen (Andini, 2017). *Pseudo F* dapat dihitung melalui Persamaan 19 (Wang & Xu, 2019).

$$Pseudo F = \frac{N-k}{k-1} \times \frac{SSB}{SSW} \quad (19)$$

dengan k menyatakan banyak cluster, N menyatakan banyak data, SSW yaitu *Sum of Square Within-cluster* dan SSB yaitu *Sum of Square Between-cluster*. Untuk nilai SSB dan SSW dapat dicari melalui Persamaan 20 dan Persamaan 21.

$$SSW = \sum_{l=1}^k \sum_{x_l \in C_1} (x_l - C_k) (x_l - \underline{x}_l)^T \quad (20)$$

$$SSB = \sum_{l=1}^k N_l (x_l - \underline{x}) (x_l - \underline{x})^T \quad (21)$$

dengan C_l menyatakan *cluster* ke- l , x_i menyatakan titik ke- i pada *cluster* ke- l , C_k menyatakan *cluster* ke- k , \underline{x}_l menyatakan titik pusat *cluster* ke- l , N_l yaitu jumlah titik pada *cluster* ke- l . Untuk $\underline{x}_l = \frac{1}{N_l} \sum_{x_i \in C_l} x_i$ dan untuk $\underline{x} = \frac{1}{N} \sum_{i=1}^N x_i$. Nilai *Pseudo F* yang lebih tinggi berkaitan terhadap *cluster* yang lebih jelas. (Wang *et al.*, 2019).

3. METODE PENELITIAN

Terdapat 210 data yang digunakan untuk melakukan *clustering* terhadap sistem ketahanan pangan di Kabupaten Pati. Data tersebut berasal dari Dinas Ketahanan Pangan Kabupaten Pati dan mencakup 21 kecamatan di wilayah tersebut. Variabel independen yang digunakan meliputi luas lahan tanam, produksi pangan, jumlah sarana dan prasarana penyedia pangan, jumlah penduduk tingkat kesejahteraan rendah, desa tanpa akses penghubung memadai, jumlah rumah tangga tanpa akses air bersih, luas wilayah, jumlah penduduk, rumah rumah tangga, dan tingkat penduduk. Sedangkan variabel dependen adalah Kecamatan yang ada di Kabupaten Pati. Data yang diperoleh kemudian diolah menggunakan *GUI Python*.

Langkah-langkah analisis yang dilakukan, yaitu mendeskripsikan data sistem ketahanan pangan menggunakan statistika deskriptif, melakukan standardisasi data, melakukan uji asumsi *representativeness of the sample*, melakukan uji asumsi non multikolinieritas, melakukan *principal component analysis* (PCA) bila asumsi non multikolinieritas tidak terpenuhi, melakukan pengelompokan *Hierarchical Clustering* menggunakan metode *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward's Method*, menentukan jumlah *cluster* optimal dengan *Calinski Harabasz Pseudo F-Statistic*, kemudian menarik kesimpulan.

4. HASIL DAN PEMBAHASAN

Penelitian ini berdasarkan data sistem ketahanan pangan yang memiliki 10 variabel dengan 9 variabel independen dan 1 variabel dependen. Pada Tabel 1 menunjukkan statistika deskriptif untuk data ketahanan pangan Kabupaten Pati tahun 2021. Statistika deskriptif adalah metode statistika yang dipakai untuk mendeskripsikan data secara global atau memberikan gambaran secara umum tentang karakteristik data yang digunakan. Pada variabel luas lahan tanam Kabupaten Pati yang memiliki 21 kecamatan sebesar 3.923,149 ha, dengan nilai *mean* dari produksi pangan sebesar 43.969,381 ton. Untuk nilai standar deviasi pada variabel sarana dan prasarana sebesar 2.775,876 yang mana nilai tersebut lebih besar daripada nilai *mean*-nya yaitu 1.595,286. Hal tersebut dapat diartikan bahwa data pada variabel tersebut semakin menyebar dari nilai *mean*-nya atau data pada variabel bersifat heterogen. Nilai *mean* dari jumlah penduduk dengan tingkat kesejahteraan rendah sebesar 2.219 atau sebesar 4% dari rata-rata jumlah penduduk yang ada di Kabupaten Pati kurang mampu untuk membeli pangan yang cukup dan bergizi.

Dari data ketahanan pangan tersebut memiliki satuan yang berbeda, sehingga dilakukan standardisasi menjadi nilai *Z-score* dengan menggunakan *GUI Python* yang akan digunakan untuk melakukan analisis *cluster*.

Tabel 1. Statistika Deskriptif Data Ketahanan Pangan Kabupaten Pati

Variabel	Mean	Std. Deviasi
Luas Lahan Tanam	3.923,149	2.336,843

Produksi Pangan	43.969,381	59.765,335
Sarana dan Prasarana Penyedia Pangan	1.595,286	2.775,876
Jumlah Penduduk Tingkat kesejahteraan Rendah	2.219	752,259
Kecamatan Tanpa Akses Penghubung Memadai	19,3	5,407
Jumlah Rumah Tangga Tanpa Akses Air Bersih	4.259,286	1.297,146
Luas Wilayah	70,146	26,030
Jumlah Penduduk	62.106	21.128,884
Jumlah Rumah Tangga	23.060,810	6.340,433
Tingkat Kepadatan Penduduk	3.900.345,03	2.355.338,15

Uji asumsi *representativeness of the sample* dapat dilakukan dengan uji *Kaiser Mayer Olkin* (KMO). Sampel dianggap mewakili seluruh populasi apabila nilai KMO $> 0,5$. Pada pengolahan menggunakan *GUI Python* diperoleh nilai KMO sebesar $0,538 > 0,5$ maka dapat dikatakan bahwa sampel yang dipakai telah mewakili populasi sehingga bisa dilanjutkan ke uji asumsi non multikolinieritas.

Uji asumsi non multikolinieritas bisa dideteksi dengan bantuan nilai *Variance Inflation Factor* (VIF) masing-masing variabel. Uji asumsi ini terpenuhi jika nilai VIF < 10 . Nilai VIF masing-masing variabel disajikan pada Tabel 2.

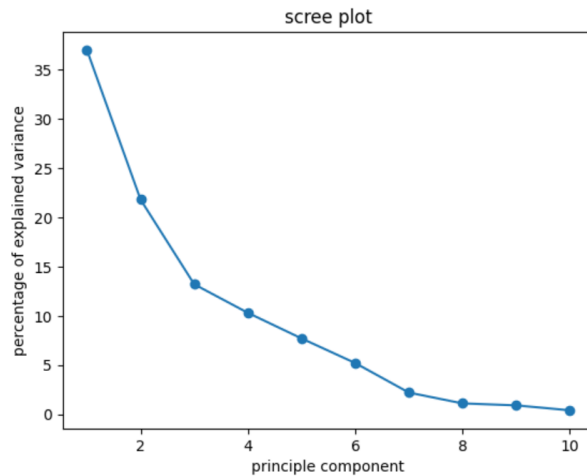
Tabel 2. Nilai VIF Variabel Ketahanan Pangan

Variabel	Nilai VIF
Luas Lahan Pertanian	18,703631
Sarana dan Prasarana Penyedia Pangan	5,476067
Jumlah Penduduk Tingkat kesejahteraan Rendah	1,719636
Kecamatan Tanpa Akses Penghubung Memadai	16,144840
Jumlah Rumah Tangga Tanpa Akses Air Bersih	29,565328
Jumlah Tenaga Kesehatan	10,612331
Luas Wilayah	40,461110
Jumlah Penduduk	54,920362
Jumlah Rumah Tangga	124,483215
Tingkat Kepadatan Penduduk	21,431554

Berdasarkan Tabel 2 diperoleh hasil bahwa terdapat 8 variabel yang memiliki nilai VIF > 10 maka terjadi multikolinieritas. Sehingga dilakukan penanganan efek multikolinieritas dengan penerapan *Principal Component Analysis* (PCA).

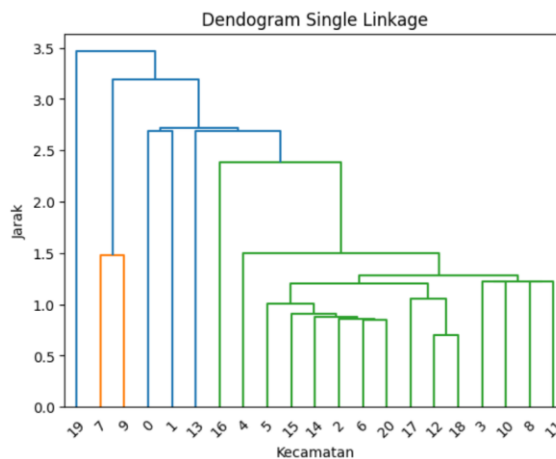
Dari proses PCA dengan membentuk matriks data dan kemudian dijadikan matriks korelasi sebagai matriks inputan, dari Persamaan 11 diperoleh nilai *eigen value* yaitu, $\lambda_1 = 3.70$, $\lambda_2 = 2.184$, $\lambda_3 = 1.32$, $\lambda_4 = 1.03$, $\lambda_5 = 0.78$, $\lambda_6 = 0.52$, $\lambda_7 = 0.23$, $\lambda_8 = 0.04$, $\lambda_9 = 0.10$, dan $\lambda_{10} = 0.11$. Kemudian, untuk menentukan jumlah komponen utama dapat dilihat melalui *scree plot*. Berdasarkan Gambar 1, terlihat bahwa patahan berbentuk siku (*elbow*) terbentuk pada nilai *component* ke-3 dan plot akan meluruh setelah *component* ke-3.

Namun, berdasarkan pada pengolahan *GUI Python* diperoleh nilai proporsi kumulatif varian yang lebih dari 80% berada pada *component* ke-4 yaitu sebesar 0,8237 atau 82,37%. Maka komponen utama yang terbentuk adalah 4 komponen. *GUI Python* pada penelitian ini telah di-*setting* secara otomatis untuk memperoleh hasil *score* komponen utama, sehingga dapat langsung digunakan ke tahap *clustering*



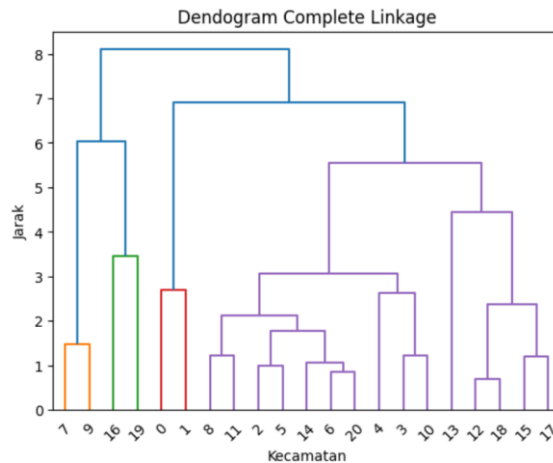
Gambar 2. Scree Plot

Pada Gambar 3 menunjukkan dendrogram *single linkage*, apabila dibentuk sejumlah 2 *cluster* maka *cluster* 1 beranggotakan kecamatan yang berada pada urutan ke-7 dan ke-9, sedangkan *cluster* 2 beranggotakan kecamatan yang berada pada urutan ke-7 dan ke-9. Jika dibentuk 3 *cluster* maka *cluster* 1 beranggotakan kecamatan yang berada pada urutan ke-19, *cluster* 2 beranggotakan kecamatan yang berada pada urutan ke-7 dan ke-9, dan *cluster* ke-3 beranggotakan kecamatan yang berada pada urutan ke-0 hingga urutan ke-11. Begitu seterusnya dapat diinterpretasikan dengan jumlah *cluster* yang diinginkan.



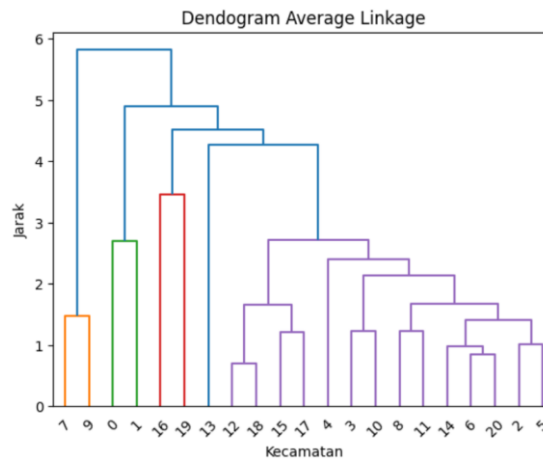
Gambar 3. Dendrogram Clustering Wilayah Ketahanan Pangan dengan Metode *Single Linkage*

Pada Gambar 4 disajikan dendrogram *complete linkage*, ketika dibentuk sebanyak 4 *cluster*, maka *cluster* 1 beranggotakan kecamatan yang berada pada urutan ke-7 dan ke-9, *cluster* 2 memiliki anggota kecamatan yang berada di urutan ke-16 dan urutan ke-19, *cluster* 3 beranggotakan kecamatan yang berada pada urutan ke-0 hingga urutan ke-1, dan *cluster* 4 beranggotakan kecamatan yang berada pada urutan ke-8 hingga urutan ke-17. Pada Gambar 4 terlihat bahwa masing-masing cluster telah ditunjukkan dengan warna yang sama sebagai arti bahwa dari masing-masing kecamatan tersebut memiliki jarak paling dekat



Gambar 4. Dendrogram *Clustering* Wilayah Ketahanan Pangan dengan Metode *Complete Linkage*

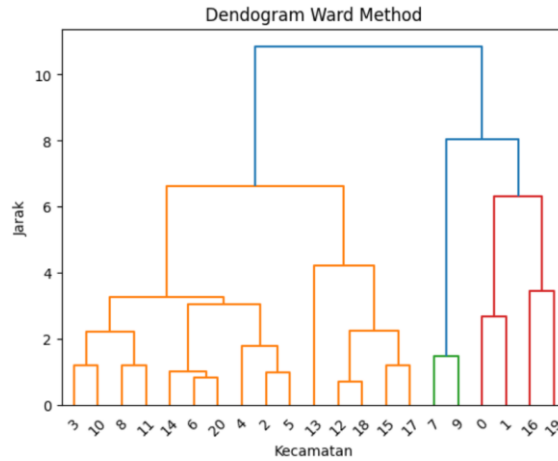
Pada dendrogram *average linkage* yang terdapat pada Gambar 5 menunjukkan ketika dibentuk sebanyak 5 *cluster*, pada *cluster* 4 hanya terdapat kecamatan urutan ke-13. Hal ini mengartikan bahwa dengan menggunakan metode *average linkage*, kecamatan ke-13 tidak mempunyai karakteristik yang sama dengan kecamatan lain. Sedangkan anggota *cluster* 1 terdiri dari kecamatan yang berada pada urutan ke-7 dan ke-9, *cluster* 2 beranggotakan kecamatan yang berada pada urutan ke-0 dan ke-1, *cluster* 3 beranggotakan kecamatan yang berada pada urutan ke-16 dan ke-19. Serta pada *cluster* 5 memiliki anggota kecamatan pada urutan ke-12 sampai urutan ke-5



Gambar 5. Dendrogram *Clustering* Wilayah Ketahanan Pangan dengan Metode *Average Linkage*

Gambar 6 menunjukkan dendrogram metode *ward*, Apabila dibentuk 2 *cluster*, maka *cluster* 1 beranggotakan kecamatan yang berada pada urutan ke-7 dan ke-9, sedangkan *cluster* 2 beranggotakan kecamatan yang berada pada urutan ke-12 sampai dengan urutan ke-1. Jika dibentuk sejumlah 3 *cluster*, maka *cluster* 1 memiliki anggota kecamatan yang berada di urutan ke-7 dan urutan ke-9, *cluster* 2 beranggotakan kecamatan yang berada pada urutan ke-12 sampai

dengan urutan ke-11, dan *cluster* 3 memiliki anggota kecamatan yang berada pada urutan ke-16 hingga urutan ke-1. Begitu seterusnya dapat diinterpretasikan sesuai dengan jumlah *cluster* yang diinginkan.



Gambar 6. Dendrogram *Clustering* Wilayah Ketahanan Pangan dengan *Ward's Method*

Dari hasil *clustering* diatas, untuk menentukan jumlah *cluster* optimumnya dilakukan dengan metode *Calinski Harabasz Pseudo F-Statistic*. Semakin besar nilai *Pseudo F*, semakin optimal pembentukan *clusternya*. Berikut disajikan nilai *Pseudo F* dari masing-masing metode pada Tabel 3.

Tabel 3. Nilai *Pseudo-F*

Jumlah Cluster	Pseudo F			
	Single Linkage	Complete Linkage	Average Linkage	Ward's Method
2	2,4565	8,9502	7,9964	9,7561
3	6,4977	10,1229	8,5772	9,9842
4	9,1312	9,9660	10,1229	10,6670
5	9,4607	13,2670	10,7808	13,2607
6	7,8592	13,6463	9,7151	13,6463
7	8,5173	14,3771	13,6991	13,6991
8	8,1020	13,6991	13,0431	14,3749

Dari Tabel 3 terlihat bahwa nilai *Pseudo-F* tertinggi sebesar 14,3711 yang merupakan nilai *Pseudo F* pada metode *Complete Linkage* dengan jumlah *cluster* 7. Hal tersebut mengartikan bahwa metode *Complete Linkage* dengan jumlah *cluster* 7 merupakan jumlah *cluster* optimal yang terbentuk dalam pengelompokan atau *clustering* menggunakan metode *Hierarchical Agglomerative Clustering*.

Apabila dilihat dari rata-rata sistem ketahanan pangan tiap *cluster* pada Tabel 4 nilai tertinggi dari rata-rata tiap *cluster* berada pada *cluster* 4 maka dapat disimpulkan bahwa *cluster* 4 merupakan *cluster* dari kumpulan kecamatan yang memiliki kualitas sistem ketahanan pangan terbaik di Kabupaten Pati . Untuk *cluster* 1 merupakan *cluster* yang memiliki kualitas ketahanan pangan yang rendah di Kabupaten Pati dikarenakan nilai dari beberapa variabel pada rata-rata *cluster* 1 lebih rendah dibandingkan dengan *cluster* yang lain. Disamping itu, wilayah yang terdapat pada *cluster* 1 merupakan kecamatan yang memiliki daerah di dataran rendah atau di daerah pesisir, sehingga kebanyakan wilayahnya terdapat tambak jadi luas lahan tanam minim.

Tabel 4. Rata-rata Sistem Ketahanan Pangan Tiap Cluster

Variabel	Cluster			
	1	2	3	4
Luas Lahan Pertanian	3847,040	2620,250	2224,245	5549,513
Produksi Pangan	28427,5	92988,33	35845,33	117994
Sarana dan Prasarana Penyedia Pangan	791,3333	1144	1323,667	1876,5
Jumlah Penduduk Tingkat kesejahteraan Rendah	<i>205,1000</i>	2974,5	2912,667	1845,5
Kecamatan Tanpa Akses Penghubung Memadai	29	20	<i>14,66667</i>	<i>14,66667</i>
Jumlah Rumah Tangga Tanpa Akses Air Bersih	5365,667	3264	4921,333	3058,5
Luas Wilayah	83,2933	49,490	75,49	50,530
Jumlah Penduduk	48507	70023,5	57859,33	103887,5
Jumlah Rumah Tangga	17843,33	24689	25975,67	36563,5
Tingkat Kepadatan Penduduk	<i>1047802</i>	3719812,015	2542176	9572915,435

(Lanjutan)Tabel 5. Rata-rata Sistem Ketahanan Pangan Tiap Cluster

Variabel	Cluster			
	5	6	7	8
Luas Lahan Pertanian	4506,7736	4445,61	3348,09	<i>1940,78</i>
Produksi Pangan	22148,4	10947,33	<i>4080,5</i>	92268
Sarana dan Prasarana Penyedia Pangan	829,6	5069,333	642	475
Jumlah Penduduk Tingkat kesejahteraan Rendah	2224,4	1697	2166,5	1523
Kecamatan Tanpa Akses Penghubung Memadai	18,4	25	19	15
Jumlah Rumah Tangga Tanpa Akses Air Bersih	4016,2	3863,333	5160	3948
Luas Wilayah	71,752	66,8	87,645	62,23
Jumlah Penduduk	58875,4	66655,33	521295	38703
Jumlah Rumah Tangga	21062,4	24123	19244,4	<i>14145</i>
Tingkat Kepadatan Penduduk	1775861,61	2984993	2133657,63	1448153,73

Keterangan: **Bold** mengartikan nilai tertinggi, *italic* nilai terendah

5. KESIMPULAN

Berdasarkan analisis dan pembahasan, dapat disimpulkan bahwa *clustering* ketahanan pangan di Kabupaten Pati dengan menggunakan metode *Agglomerative Hierarchical Clustering* dan pengoptimalan jumlah *cluster* menggunakan *Calinski Harabasz Pseudo F-Statistic*, diperoleh jumlah *cluster* optimal sebanyak 7 *cluster* dengan teknik penghitungan jarak menggunakan metode *Complete Linkage*. Manfaat dari penelitian ini menggunakan algoritma *clustering* yaitu, lebih mempermudah dalam mengidentifikasi karakteristik antar wilayah ketahanan pangan yang memiliki kesamaan paling dekat dan dapat mengidentifikasi kecamatan yang memiliki sistem ketahanan pangan tinggi maupun sistem ketahanan pangannya rendah. Berdasarkan nilai rata-rata tiap *cluster* diperoleh bahwa *cluster* 4 merupakan *cluster* dari kumpulan kecamatan yang memiliki kualitas ketahanan pangan terbaik di Kabupaten Pati dan *cluster* 1 merupakan *cluster* yang memiliki kualitas ketahanan pangan yang rendah di Kabupaten Pati dikarenakan nilai dari semua variabel pada rata-rata *cluster* 1 lebih rendah dibandingkan dengan *cluster* yang lain. Dengan demikian, dapat membantu Pemerintah Kabupaten Pati dan Dinas Ketahanan Pangan untuk melakukan kebijakan terhadap wilayah yang memiliki sistem ketahanan pangan rendah untuk meningkatkan ketahanan di wilayah tersebut. Serta dengan menggunakan *GUI Python* lebih mempermudah dalam proses analisis *clustering*, baik dalam pengolahan data maupun analisis datanya.

DAFTAR PUSTAKA

- Andini, P. D. T. 2017. Pengelompokan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Pendidikan Formal Wajib Belajar 12 Tahun Menggunakan Cluster Hierarchy. *Skripsi*. Institut Teknologi Sepuluh November.
- Anton, H., Rorres, C. 2004. *Aljabar Linear Algebra: Versi Aplikasi*. Edisi 8. Diterjemahkan oleh: Indriasari, R., & Harmein, I. Jakarta: Erlangga. Terjemahan dari: *Elementary Linear Algebra: Application Version*.
- Baliwati, Y.F., A. Khomsan, C. M. Dwiriani. 2004. *Pengantar Pangan dan Gizi*. Penebar Swadaya: Jakarta.
- Han, J., Kamber, M., Pei, J. 2012. *Data Mining Concepts and Techniques Third Edition*. San Massachusetts (US): Morgan Kaufmann Publisher.
- Johnson, R., Winchern, D. 2007. *Applied Multivariate Statistical Analysis* (6nd Edition ed.). United State of America: Pearson Prentice-Hall.
- Kaur, Maninderjit, Sushil, K. G. 2014. Survey on Clustering Techniques in Data Mining for Software Engineering. *International Journal of Advanced and Innovative Research*, No. 3, Hal: 238-243.
- Kingrani, Kumr S., Levene, Mark, Zhang, Dell. 2018. Estimating the Number of Cluster Using Diversity. *Artificial Intelligence Research* Vol. 7, No. 1, Hal: 15-22. ISSN 1927-6974.
- Larose, Daniel, T., Chantal D. 2015. *Data Mining and Predictive Analytics. Second Edition*, John Wiley & Sons
- Nugroho, S. 2008. *Statistika Multivariat Terapan*. Edisi Pertama. Bengkulu: UNIP Press.
- Sudirman, Windarto, A.P., Wanto, A. 2018. Data Mining Tools | rapidminer: K-means Method on Clustering of Rice Crops in Indonesia. *IOP Conferences Series: Materials Science and Engineering* Vol. 420, No. 012089. DOI:10.1088/1757-899X/420/1/012089.
- Suryana, A. 2014. Menuju Ketahanan Pangan Indonesia Berkelanjutan 2025: Tantangan dan Penanganannya. *Agro Ekonomi* Vol. 23, No. 2, Hal: 123-135.
- Santoso, S. 2018. *Mahir Statistik Multivariat dengan SPSS*. Jakarta: PT Elex Media Komputindo.
- Wang, X., Xu, Y., 2019. An Improved Index for Clustering Validation Based on Silhouette Index and Calinski-Harabasz Index. *IOP Conf. Series: Materials Science and Engineering* Vol. 569, No. 052024. DOI:10.1088/1757-899X/569/5/052014.