

PENGEROMBOLAN KABUPATEN/KOTA DI SUMATERA UTARA BERDASARKAN PRODUKTIVITAS TANAMAN PANGAN

Regency/City Clustering in North Sumatra Based on Food Crop Productivity

Junjun Wijaya*, Monika Stevany Manurung**
*BPS Kabupaten Dairi, *E-mail*: junjun.wijaya@bps.go.id
**BPS Kabupaten Dairi

ABSTRAK

Pertanian merupakan sektor andalan yang mendorong perekonomian di kabupaten/kota. Tanaman pangan merupakan subsektor dari pertanian dengan kontribusi paling besar dalam PDRB Provinsi Sumatera Utara tahun 2021. Untuk memaksimalkan potensi tanaman pangan, maka dibutuhkan pengidentifikasian dari sebaran produktivitas. Dengan demikian, pemerintah dapat mengambil kebijakan dalam meningkatkan potensi subsektor tanaman pangan di masing-masing wilayah. Tujuan dari penelitian ini adalah menggerombolkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan produktivitas tanaman pangan dengan menggunakan metode gerombol terbaik. Penelitian ini menggunakan metode hirarki dan non hirarki dengan uji validitas internal yaitu indeks *Connectivity*, indeks *Silhouette*, dan indeks *Dunn* sehingga diperoleh metode penelitian terbaik adalah metode *K-Means Clustering* dengan jumlah optimum *cluster* yang dihasilkan sebanyak 2 *cluster*. Hasil penggerombolan adalah *cluster* 1 dengan produktivitas tanaman pangan rendah, sedangkan *cluster* 2 dengan produktivitas tanaman pangan tinggi.

Kata kunci: Tanaman Pangan, Produktivitas, *K-Means Clustering*

ABSTRACT

Agriculture is the mainstay sector that drives the economy in districts/cities. Food crops are the subsector of agriculture with the largest contribution to the GRDP of North Sumatra Province in 2021. To maximize the potential of food crops, it is necessary to identify the distribution of productivity. Thus, the government can adopt policies to increase the potential of the food crop sub-sector in each region. The aim of this research is to cluster districts/cities in North Sumatra Province based on food crop productivity using the best cluster method. This research uses hierarchical and non-hierarchical methods with internal validity tests, namely the Connectivity index, Silhouette index, and Dunn index so that the best research method is obtained, namely the K-Means Clustering method with the optimum number of clusters produced being 2 clusters. The result of clustering is cluster 1 with low crop productivity, while cluster 2 has high crop productivity.

Keywords: Food Crops, Productivity, K-Means Clustering

I. PENDAHULUAN

Sektor pertanian merupakan sektor yang berperan strategis dalam pembangunan nasional. Sektor pertanian menjadi sektor ketiga terbesar yang mendominasi pertumbuhan ekonomi setelah sektor industri dan perdagangan pada Produk Domestik Bruto (PDB) Indonesia tahun 2022. Sumatera Utara merupakan salah satu daerah dengan potensi sektor pertanian yang melimpah di pulau sumatera. Sektor pertanian merupakan sektor dengan kontribusi paling besar dalam menopang perekonomian di provinsi Sumatera Utara. Berdasarkan data Produk Domestik Regional Bruto (PDRB) Lapangan Usaha Provinsi Sumatera Utara tahun 2021, sektor pertanian menyumbang sebesar 22,04 persen terhadap perekonomian provinsi (Badan Pusat Statistik, 2021).

Sektor pertanian terdiri dari lima subsektor antara lain tanaman pangan, hortikultura, perkebunan, peternakan serta jasa pertanian dan perburuan. Subsektor tanaman pangan terdiri dari padi (padi sawah dan padi ladang) dan palawija (jagung, kacang tanah, kacang hijau, kacang kedelai, ubi jalar, dan ubi kayu). Tanaman pangan merupakan subsektor dengan kontribusi paling besar terhadap perekonomian di antara subsektor lainnya. Dengan kata lain, tanaman pangan merupakan subsektor yang paling dominan. Hal tersebut dikarenakan tanaman pangan berperan penting dalam penyediaan makanan pokok bagi masyarakat terutama tanaman pangan padi.

Luas panen padi Provinsi Sumatera Utara tahun 2021 sebesar 385.405 hektar dengan produksi sebesar 2.004.143 ton Gabah Kering Giling (GKG). Apabila dibandingkan dengan tahun 2020, luas panen dan produksi tersebut mengalami penurunan (BPS, 2021). Kabupaten/kota dengan produksi padi tertinggi adalah Kabupaten Deli Serdang, Serdang Bedagai, dan Simalungun.

Potensi tanaman pangan di Sumatera Utara perlu dipertahankan dan dikelola agar ketersediaan pangan tetap terjaga. Untuk itu, sebaran tanaman pangan di Sumatera Utara perlu diidentifikasi berdasarkan produktivitas kabupaten/kota. Gambaran sebaran tersebut akan membantu pemerintah untuk mengidentifikasi kelompok pemetaan dan mengambil kebijakan dalam meningkatkan potensi subsektor tanaman pangan di masing-masing wilayah. Pengelompokan kabupaten/kota di Sumatera Utara berdasarkan sebaran tanaman pangan dapat dilakukan dengan analisis gerombol (*clustering*). Hal ini didukung oleh penelitian yang dilakukan oleh Safitri, dkk (2018) yang berjudul penggerombolan provinsi di Indonesia berdasarkan produktivitas tanaman pangan menggunakan metode *K-Error*. Penelitian tersebut menghasilkan sebanyak 7 gerombol dengan kategori produktivitas dari sangat rendah hingga sangat tinggi. Gerombol 2 dan 3 merupakan gerombol dengan kategori sangat rendah, gerombol 7 termasuk kategori rendah. Sedangkan kategori sedang terdapat pada gerombol gerombol 1 dan 6, sementara kategori tinggi dan sangat tinggi masuk dalam kategori 4 dan 5.

Analisis gerombol (*clustering*) adalah salah satu teknik analisis pengelompokan objek yang memiliki kemiripan menjadi beberapa gerombol tertentu. Objek di dalam satu gerombol akan memiliki kemiripan karakteristik yang tinggi sedangkan objek di luar gerombol tersebut akan memiliki tingkat kemiripan yang rendah. Dengan kata lain, tingkat variasi dalam gerombol menjadi rendah sedangkan antar gerombol menjadi tinggi (Mattjik dan Sumertajaya, 2011). Secara garis besar terdapat dua metode pada analisis gerombol (*clustering*) diantaranya metode hirarki dan non hirarki. Salah satu metode hirarki adalah *average linkage*, sedangkan metode non hirarki adalah *K-Means clustering*. Metode *average linkage* menghitung jarak dua gerombol yang disebut sebagai jarak rata-rata. Berdasarkan penelitian Alfina, dkk (2012) menyatakan bahwa metode tersebut merupakan metode *clustering* paling baik yang dihasilkan. Sedangkan metode *K-Means* mempartisi X dalam k buah gerombol dari sekumpulan objek $X = X_1, X_2, X_3, \dots, X_n$ (Johnson dan Wichern, 2007).

Penelitian ini menggunakan variabel produktivitas dalam melakukan penggerombolan kabupaten/kota di Provinsi Sumatera Utara. Produktivitas komoditas pertanian dapat menggambarkan situasi ketahanan dan kerentanan pangan wilayah kabupaten/kota. Selain itu, penelitian ini menggunakan data terbaru yaitu data tahun 2010 hingga 2020 sehingga hasil yang diperoleh diharapkan dapat menggambarkan kondisi pertanian saat ini.

Berdasarkan rumusan tersebut, maka tujuan dari penelitian ini adalah menggerombolkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan produktivitas dari komoditas tanaman pangan tahun 2010-2020. Untuk menjawab tujuan tersebut penggerombolan akan menggunakan metode terbaik di antara *average linkage* dan *K-Means clustering*. Hasil analisis gerombol (*clustering*) kedua metode tersebut akan diperbandingkan berdasarkan evaluasi model. Dengan demikian, dapat diidentifikasi daerah dengan produktivitas tanaman pangan potensial yang dapat dioptimalkan.

II. TINJAUAN PUSTAKA

A. Analisis Gerombol (*Clustering*)

Analisis gerombol (*clustering*) merupakan salah satu teknik analisis multivariat metode interdependen. Tujuan dari analisis gerombol adalah mengelompokkan objek berdasarkan kemiripan karakteristik tertentu ke dalam gerombol yang sama (Mattjik dan Sumertajaya, 2011). Dengan demikian, satu gerombol akan terdiri atas objek dengan karakteristik homogen, sedangkan objek antar gerombol memiliki karakteristik heterogen. Pendekatan yang digunakan untuk mengukur kemiripan antar objek dinyatakan dalam jarak (*distance*). Pasangan objek dengan jarak yang pendek akan cenderung lebih mirip sehingga dikelompokkan pada gerombol yang sama. Pengukuran jarak tersebut dilakukan dengan mempertimbangkan sifat variabel, skala pengukuran, dan karakteristik masalah yang diteliti.

Jarak (*distance*)

Jarak merupakan ukuran umum yang digunakan dalam mengukur kemiripan antar objek. Jarak yang digunakan apabila antar variabel saling bebas atau tidak terdapat multikolinearitas pada data adalah jarak *euclidean*. Jarak *euclidean* merupakan akar jumlah kuadrat perbedaan nilai objek tiap variabel. Sementara itu, apabila terdapat korelasi pada data digunakan jarak *Mahalanobis*.

B. Metode Cluster

Metode Hierarki

Pengukuran kemiripan observasi dilakukan secara bertahap yang terdiri dari:

(a) *Agglomerative nested* (Pemusatan)

Menurut (Johnson, 1967), metode hirarki dilakukan dengan cara yaitu terdapat sekumpulan N objek dan sebuah matrik berukuran N x N kemudian yaitu

1. Bentuk gerombol sebanyak N dengan masing-masing gerombol memiliki sebuah objek.
2. Cari sepasang gerombol yang jaraknya terdekat, dan dijadikan sebuah gerombol baru. Maka akan terbentuk sebanyak N – 1 gerombol.
3. Hitung jarak antar gerombol yang baru dengan masing-masing gerombol yang lainnya.

Ulangi tahapan pada poin 2 dan 3 hingga semua objek menjadi sebuah gerombol dengan N objek. Apabila N objek yang dikelompokkan membentuk satu gerombol besar maka artinya tidak ada gunanya dilakukan penggerombolan. Setiap objek diasumsikan sebagai satu gerombol tersendiri. Kemudian, diukur jarak antar gerombol yang berpeluang terjadi pengelompokkan dan begitu seterusnya. *Agglomerative nested* terdiri dari *single linkage*, *complete linkage*, dan *average linkage*.

Metode *Single Linkage* merupakan metode hirarki yang paling sederhana, dimana gerombol dibentuk dari entitas objek dengan menggabungkan objek terdekat, di mana istilah objek terdekat berarti jarak terkecil atau *similarities* terbesar. *Complete Linkage* berlangsung dengan cara yang hampir sama seperti *Single Linkage*, dengan satu pengecualian penting, yaitu pada setiap tahap. *Complete Linkage* memastikan bahwa semua objek dalam sebuah gerombol berada dalam jarak maksimum (*similarities* minimum) satu sama lain. Sementara itu, metode *Average Linkage* menghitung jarak rata-rata dua gerombol di mana jarak tersebut dihitung pada masing-masing gerombol dengan persamaan berikut (Johnson & Wichern, 2007):

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)}N_W} \tag{1}$$

Keterangan:

d_{ik} : Jarak antar objek *i* dalam gerombol (*UV*) dan objek *k* dalam gerombol *W*

$N_{(UV)}$ dan N_w : Jumlah objek dalam gerombol (UV) dan (W)

Metode *centroid* merupakan rata-rata dari semua objek dalam gerombol. Jarak antar *centroid* adalah jarak antar gerombol. Sedangkan, metode *Ward* didasari oleh hilangnya informasi yang disebabkan oleh penggabungan objek menjadi gerombol. Metode ini diukur dengan SSE (*Sum Square Error*) sebagai fungsi objektif (Johnson & Wichern, 2007). Tujuan metode *Ward* meminimumkan ragam dalam cluster dan memaksimalkan ragam antar objek di gerombol lainnya. Nilai SSE (*Sum Square Error*) untuk gerombol dengan anggota satu objek atau satu item saja adalah bernilai nol.

(b) *Divisive analysis* (Penyebaran)

Semua observasi dianggap menjadi satu gerombol besar. Observasi dengan nilai kemiripan yang rendah akan dipisahkan dan begitu seterusnya.

Metode Non Hierarki

Metode ini dilakukan dengan membentuk *initial cluster* pusat. Objek dengan jarak tertentu dikelompokkan pada gerombol yang terbentuk. Objek dapat berpindah mempertimbangkan ukuran jarak dengan gerombol lain. Metode non hierarki ini biasa disebut dengan *K-Means Clustering*.

K-Means Clustering merupakan salah satu metode analisis gerombol non hirarki. Metode ini dapat digunakan mempartisi objek ke dalam kelompok-kelompok berdasarkan kedekatan karakteristik. Dengan demikian, objek dengan karakteristik yang sama dikelompokkan dalam satu gerombol yang sama sedangkan objek dengan karakteristik berbeda dikelompokkan ke dalam gerombol lain. Tujuan pengelompokan adalah untuk meminimalkan *objective function* yang diset dalam proses pengelompokan, yang pada dasarnya berusaha untuk meminimalkan variasi dalam satu klaster dan memaksimalkan variasi antar klaster.

Jika diberikan sekumpulan objek $X = X_1, X_2, X_3, \dots, X_n$ maka algoritma *K-Means Clustering* akan mempartisi X dalam k buah gerombol, di mana setiap gerombol memiliki centroid dari objek-objek dalam gerombol tersebut. Pada tahap awal algoritma *K-Means Clustering* dipilih secara acak k buah objek sebagai centroid, kemudian jarak antara objek dengan centroid dihitung dengan menggunakan jarak *Euclidean*, objek ditempatkan dalam gerombol yang terdekat dihitung dari titik tengah gerombol. Centroid baru ditetapkan jika semua objek sudah ditempatkan dalam gerombol terdekat. Proses penentuan *centroid* dan penempatan objek dalam gerombol diulangi sampai nilai centroid konvergen (*centroid* dari semua gerombol tidak berubah lagi) (Johnson dan Wichern, 2007).

Evaluasi Model

(a) Indeks *Dunn*

Dunn Index merupakan salah satu metode validasi gerombol yang diajukan oleh J.C. Dunn. *Dunn Index* merupakan salah satu metode validasi hasil penggerombolan dengan mengukur diameter gerombol dan jarak antara dua gerombol sehingga diperoleh jumlah gerombol yang optimum (Khajvand dan Tarokh, 2011). Metode ini memberikan nilai terbaik untuk menghasilkan gerombol yang dalam gerombol memiliki kemiripan tinggi sedangkan antar gerombol memiliki kemiripan yang rendah. *Dunn Index* bertujuan untuk mengidentifikasi gerombol yang terpisah dengan baik dengan cara menghitung rasio jarak antar gerombol minimal dan jarak dalam gerombol maksimal. Artinya, semakin tinggi nilai *Dunn Index*, maka jumlah gerombol yang dihasilkan semakin optimal.

(b) Indeks *Connectivity*

Indeks *Connectivity* dilambangkan dengan Conn. Semakin kecil nilai Indeks *Connectivity* maka diindikasikan menunjukkan banyak gerombol yang terbentuk lebih baik atau optimal (Halim & Widodo, 2017).

(c) Indeks *Silhouette*

Indeks *Silhouette* digunakan untuk menghitung rata-rata dari titik yang ada pada himpunan data. Metode ini dapat digunakan sebagai validasi pada seluruh data, baik yang bersifat tunggal atau keseluruhan gerombol. Indeks *Silhouette* merupakan metode yang paling umum untuk memvalidasi gerombol yang menghubungkan separasi dan kohesi. Semakin besar rata-rata rentang nilai indeks *Silhouette* dari seluruh objek dalam suatu gerombol menunjukkan banyak kelompok hasil pengelompokan yang semakin baik (optimal). Evaluasi metode terbaik dapat menggunakan koefisien korelasi *Cophenetic*.

III. METODE PENELITIAN

A. Data dan Sumber Data

Data yang digunakan pada penelitian kuantitatif ini merupakan data sekunder bersumber dari BPS Provinsi Sumatera Utara. Variabel yang digunakan merupakan produktivitas dari komoditas tanaman pangan, antara lain jagung, kacang hijau, kacang kedelai, kacang tanah, ubi jalar, ubi kayu, serta padi sawah dan padi ladang. Produktivitas merupakan rasio produksi tanaman pangan terhadap luas lahan yang digunakan dalam usaha pertanian. Data tersebut diperoleh melalui Survei Ubinan. Sementara itu, objek penelitian ini adalah seluruh

kabupaten/kota di Provinsi Sumatera Utara (kecuali Kota Sibolga) dengan jumlah 32 kabupaten/kota.

B. Metode Analisis

Berdasarkan tujuan penelitian, maka metode yang digunakan pada penelitian ini adalah analisis *cluster*. Metode tersebut digunakan karena mampu mengelompokkan setiap kabupaten/kota dengan produktivitas tanaman pangan yang mirip ke dalam gerombol yang sama. Untuk itu, tahapan yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Mengidentifikasi *missing value* dan melakukan imputasi terhadap *missing value* menggunakan nilai rata-rata kabupaten/kota menurut komoditas
2. Melakukan analisis deskriptif
3. Melakukan standardisasi apabila satuan antar variabel berbeda
4. Melakukan uji asumsi non multikolinearitas
5. Membentuk matriks jarak *euclidean*
6. Penggerombolan dengan metode *cluster* hirarki
 - a. Menentukan metode terbaik di antara metode *cluster* hirarki berdasarkan koefisien *cophenetic*
 - b. Menentukan jumlah *cluster k* optimal
7. Penggerombolan dengan metode *cluster* non hirarki
 - a. Menentukan *initial cluster*
 - b. Melakukan validitas internal
8. Membandingkan hasil *cluster* metode hirarki dan non hirarki
9. Profilisasi terhadap hasil *cluster* dengan metode terbaik
10. Interpretasi

IV. HASIL, ANALISIS, DAN PEMBAHASAN

A. Analisis Deskriptif

Penelitian ini tidak melakukan standarisasi data karena setiap variabel yang digunakan dalam menggerombolkan kabupaten/kota memiliki satuan yang sama, yakni ku/ha. Sebelum

melakukan analisis deskriptif, peneliti terlebih dahulu melakukan imputasi *missing value* pada beberapa kabupaten/kota dengan nilai provinsi. Dengan demikian, peneliti menggunakan semua kabupaten/kota, kecuali Kota Sibolga untuk mempertahankan keterwakilan data agar representatif. Analisis deskriptif pada penelitian ini secara umum bertujuan menggambarkan produktivitas tanaman pangan di Provinsi Sumatera Utara. Tanaman pangan merupakan salah satu dari kebutuhan dasar masyarakat.

Tabel 1. Hasil Analisis Deskriptif

Tanaman Pangan	Min	Q1	Med	Mean	Q3	Maks
Jagung	38.90	43.57	48.14	48.83	52.29	67.09
Kacang Hijau	4.389	7.540	9.890	10.221	12.916	15.578
Kacang Kedelai	1.961	5.934	10.105	8.775	11.550	13.785
Kacang Tanah	3.779	8.694	10.470	9.729	11.071	13.440
Ubi Jalar	29.73	104.52	135.79	127.54	148.65	200.92
Ubi Kayu	174.7	233.8	278.8	278.0	318.8	401.5
Padi Sawah dan Ladang	34.97	42.69	47.00	47.52	51.75	58.30

Tabel 1 menunjukkan bahwa produktivitas tanaman pangan palawija ubi jalar dan ubi kayu memiliki range yang cukup tinggi. Hal tersebut mengindikasikan bahwa terdapat *gap* produktivitas dari kedua jenis tanaman pangan tersebut di kabupaten/kota Provinsi Sumatera Utara. Pada tabel tersebut juga dapat ditunjukkan bahwa jenis tanaman pangan dengan rata-rata produktivitas tertinggi adalah ubi kayu dan ubi jalar. Sebaliknya, jenis tanaman pangan dengan rata-rata produktivitas terendah adalah kacang kedelai dan kacang tanah.

B. Analisis Inferensia

Uji Non Multikolinearitas

Salah satu asumsi yang harus dipenuhi dalam analisis cluster adalah setiap variabel saling *orthogonal* atau tidak berkorelasi satu dengan yang lain. Uji non multikolinearitas bertujuan untuk mengidentifikasi korelasi di antara variabel. Asumsi tidak terpenuhi apabila antar variabel memiliki korelasi $> 0,8$. Dengan kata lain, terjadi multikolinearitas pada data.

Tabel 2. Hasil Uji Non Multikolinearitas

Tanaman Pangan	Jagung	Kacang Hijau	Kacang Kedelai	Kacang Tanah	Ubi Jalar	Ubi Kayu	Padi Sawah dan Ladang
Jagung	1.0000	-0.0498	0.0723	0.3296	0.1095	0.2922	0.5062
Kacang Hijau	-0.0498	1.0000	0.6072	0.2164	0.0953	0.0099	0.1779
Kacang Kedelai	0.0723	0.6072	1.0000	0.1068	-0.1213	0.1424	0.3299
Kacang Tanah	0.3296	0.21635	0.1068	1.0000	0.5213	0.5947	0.3007
Ubi Jalar	0.1097	0.0953	-0.1213	0.5213	1.0000	0.4131	0.0244
Ubi Kayu	0.2922	0.0099	0.1424	0.5947	0.4131	1.0000	0.4850
Padi Sawah dan Ladang	0.5062	0.1779	0.3299	0.3007	0.0244	0.4850	1.0000

Tabel 2 merupakan nilai korelasi antar variabel. Berdasarkan tabel tersebut dapat ditunjukkan bahwa tidak terdapat korelasi antar variabel dengan nilai $> 0,8$. Artinya, data sudah memenuhi asumsi dalam analisis *cluster*, yakni tidak terdapat multikolinearitas pada data. Dengan demikian, jarak *euclidean* dapat digunakan untuk menggerombolkan kabupaten/kota berdasarkan kemiripan produktivitas.

Hasil Pengelompokkan dengan Metode Cluster Hirarki

Penggerombolan dilakukan secara bertahap. Untuk menentukan metode terbaik di antara metode cluster hirarki, maka digunakan korelasi *Cophenetic*. Metode dengan koefisien korelasi *Cophenetic* tertinggi merupakan metode terbaik dalam menggerombolkan kabupaten/kota.

Tabel 3. Koefisien Korelasi *Cophenetic*

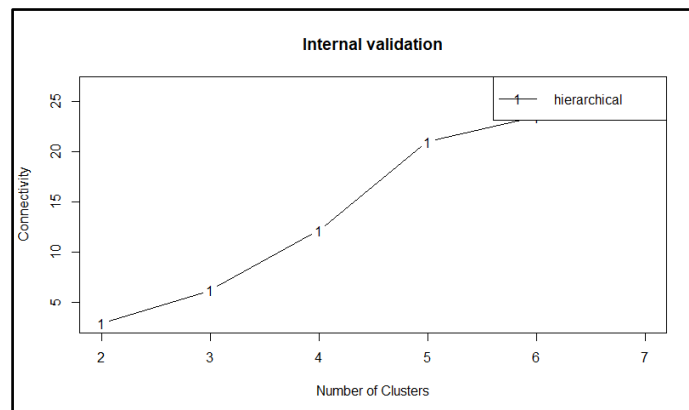
<i>Single linkage</i>	<i>Complete linkage</i>	<i>Average linkage</i>	<i>Centroid linkage</i>	<i>Ward's linkage</i>
0.5082	0.6590	0.6710	0.6486	0.6338

Berdasarkan Tabel 3 dapat ditunjukkan bahwa metode cluster hirarki dengan model terbaik adalah *average linkage*. Hal tersebut ditunjukkan dengan koefisien korelasi tertinggi dibandingkan dengan metode lainnya. Dengan demikian, penggerombolan dengan metode cluster hirarki pada penelitian ini menggunakan *average linkage*. Selanjutnya, untuk menentukan jumlah cluster optimal dilakukan tahapan evaluasi cluster dengan indeks validitas antara lain indeks *Connectivity*, indeks *Silhouette*, dan indeks *Dunn*.

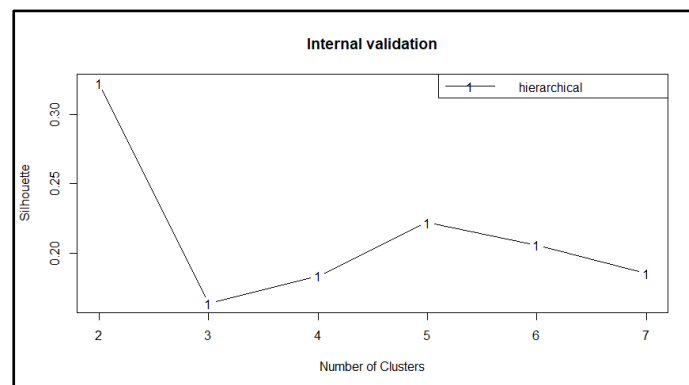
Tabel 4. Evaluasi Model

Indeks Validitas	Jumlah Cluster						
	2	3	4	5	6	7	8
<i>Connectivity</i>	7.7627	13.7329	19.0905	23.4258	25.1036	29.8401	31.8401
<i>Dunn</i>	0.2213	0.2277	0.2675	0.3281	0.3281	0.3526	0.3526
<i>Silhouette</i>	0.4279	0.3549	0.3650	0.3508	0.3379	0.3921	0.3814

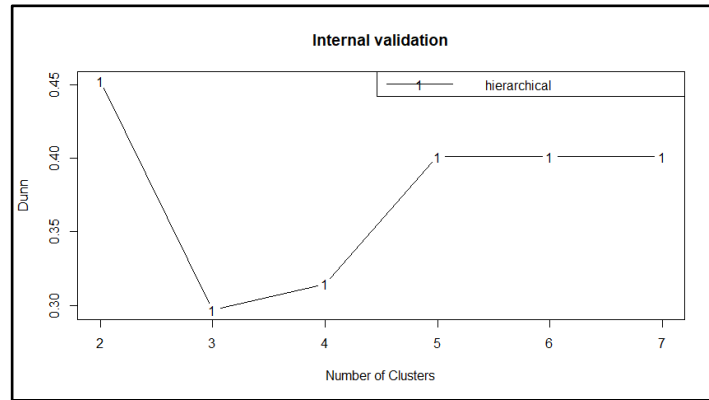
Tabel 4 merupakan hasil validasi internal yang bertujuan mengidentifikasi jumlah *cluster* optimal yang dapat dibentuk dengan metode hirarki *average linkage*. Tabel tersebut dapat digambarkan oleh Gambar 1. Berdasarkan tabel tersebut jumlah *cluster* optimal yang mampu menggerombolkan kabupaten/kota berdasarkan produktivitas adalah dua. Hal tersebut dapat ditunjukkan oleh indeks *connectivity* terendah dan indeks *silhouette* tertinggi berada pada jumlah *cluster* yang berjumlah dua dibandingkan dengan jumlah *cluster* lain. Sementara itu, indeks *dunn* menunjukkan bahwa jumlah *cluster* optimal yang dapat dibentuk adalah tujuh.



(a) Indeks *Connectivity*



(b) Indeks *Silhouette*



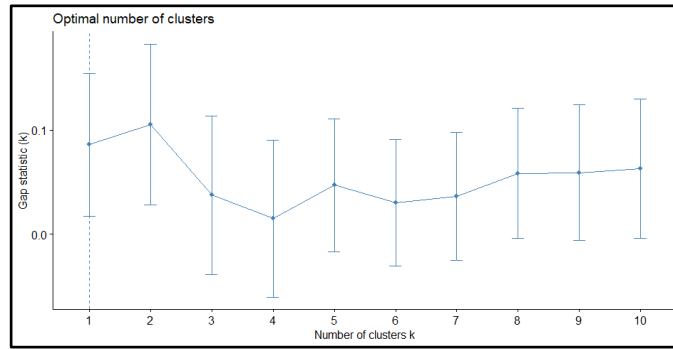
(c) Indeks *Dunn*

Gambar 1. Grafik indeks validitas internal

Dengan demikian, hasil penggerombolan kabupaten/kota berdasarkan produktivitas tanaman pangan di Provinsi Sumatera Utara dengan metode *hirarki average linkage* terdiri dari dua *cluster* (Tabel 5). Penggerombolan tersebut dapat ditunjukkan oleh Gambar 2.

Tabel 5. Hasil Penggerombolan berdasarkan Metode *Average Linkage*

Kabupaten /kota	Cluster	
	1	2
Kabupaten	Nias, Mandailing Natal, Tapanuli Tengah, Labuhan Batu, Karo, Nias Selatan, Labuhanbatu Selatan, Labuhanbatu Utara, Nias Utara, dan Nias Barat	Tapanuli Selatan, Tapanuli Utara, Toba Samosir, Asahan, Simalungun, Dairi, Deli Serdang, Langkat, Humbang Hasundutan, Pakpak Bharat, Samosir, Serdang Bedagai, Batu Bara, Padang Lawas Utara, dan Padang Lawas.
Kota	Tanjungbalai, Pematangsiantar, Tebing Tinggi, Medan, Binjai, dan Gunungsitoli.	Padang Sidempuan



(c) Metode Gap Statistik

Gambar 3. Penentuan jumlah *cluster* terbaik

Hasil metode *Elbow* Gambar 3(a) menunjukkan bahwa jumlah *cluster* optimal adalah dua. Hal tersebut ditunjukkan dengan *total within sum of square* yang turun secara signifikan pada *cluster* tersebut, yakni ditandai dengan garis yang curam atau pola siku. Jumlah *cluster* terbaik juga dapat diidentifikasi dengan koefisien *silhouette* terbesar Gambar 3(b), yakni jumlah *cluster* optimal adalah dua. Selain itu, metode *gap* statistik juga menunjukkan bahwa jumlah *cluster* yang optimal adalah dua ditunjukkan dengan garis paling tinggi.

Berdasarkan ketiga metode dalam menentukan jumlah *cluster* tersebut, maka dapat disimpulkan bahwa jumlah *cluster* optimal dalam menggerombolkan kabupaten/kota menurut produktivitas tanaman pangan di Provinsi Sumatera Utara adalah dua (Tabel 6). Penggerombolan tersebut dapat ditunjukkan oleh Gambar 4.

Tabel 6. Hasil Penggerombolan berdasarkan Metode *K-Means Clustering*

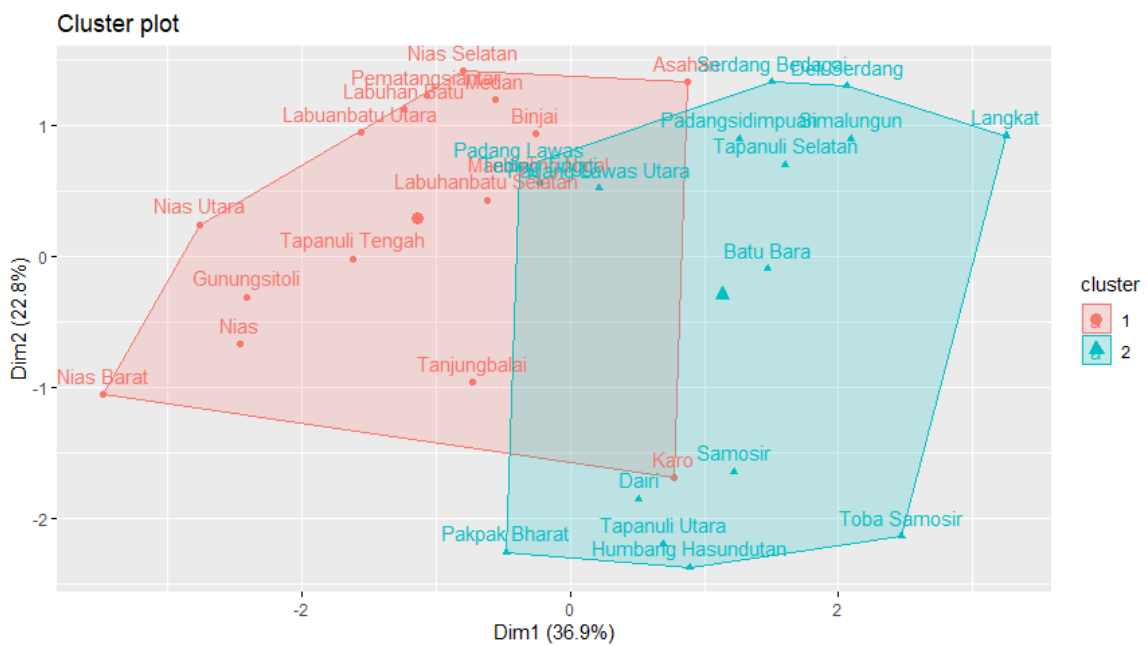
Kabupaten/kota	Cluster	
	1	2
Kabupaten	Nias, Mandailing Natal, Tapanuli Tengah, Labuhan Batu, Asahan, Karo, Nias Selatan, Labuhanbatu Selatan, Labuhanbatu Utara, Nias Utara, dan Nias Barat	Tapanuli Selatan, Tapanuli Utara, Toba Samosir, Simalungun, Dairi, Deli Serdang, Langkat, Humbang Hasundutan, Pakpak Bharat, Samosir, Serdang Bedagai, Batu Bara, Padang Lawas Utara, dan Padang Lawas.
Kota	Tanjungbalai, Pematangsiantar, Medan, Binjai, dan Gunungsitoli.	Tebing Tinggi dan Padang Sidempuan

Selanjutnya, sama halnya dengan metode hirarki, pada jumlah *cluster* non hirarki *K-Means Clustering* juga dilakukan validasi internal. Hal tersebut bertujuan menguji ketepatan dari jumlah *cluster* yang dihasilkan.

Tabel 7. Evaluasi Model

Indeks Validitas	Jumlah Cluster					
	2	3	4	5	6	7
<i>Connectivity</i>	7.7115	19.1659	24.7647	25.1111	27.9440	32.6806
<i>Dunn</i>	0.2052	0.2129	0.2249	0.3042	0.3076	0.4137
<i>Silhouette</i>	0.4548	0.3287	0.3316	0.3493	0.3403	0.4006

Berdasarkan validitas internal dengan ketiga indeks di atas terhadap kedua metode tersebut baik hirarki maupun non hirarki, maka metode terbaik yang digunakan pada penelitian ini adalah *K-Means Clustering*. Jumlah *cluster* optimal dengan metode tersebut adalah dua dengan penggerombolan kabupaten/kota pada Tabel 6. Tahapan selanjutnya adalah profilisasi dengan mengidentifikasi karakteristik dari masing-masing *cluster* yang telah terbentuk. Penelitian ini menggunakan ukuran pemusatan rata-rata seperti pada Tabel 7.



Gambar 4. Penentuan jumlah *cluster* terbaik

Tabel 8. Profilisasi

<i>Cluster</i>	Jagung	Kacang Hijau	Kacang Kedelai	Kacang Tanah	Ubi Jalar	Ubi Kayu	Padi sawah dan ladang
1	47.2512	9.9970	8.0596	8.3034	106.4426	227.4638	45.6901
2	50.4002	10.4442	9.4902	11.1551	148.6276	328.5145	49.3409

Tabel 7 dapat ditunjukkan bahwa kabupaten/kota yang termasuk ke dalam *cluster* 1 cenderung memiliki rata-rata produktivitas jagung, kacang hijau, kacang kedelai, kacang tanah, ubi jalar, ubi kayu, dan padi yang rendah dibandingkan dengan *cluster* 2. Sebaliknya, kabupaten/kota dengan produktivitas jagung, kacang hijau, kacang kedelai, kacang tanah, ubi jalar, ubi kayu, dan padi cenderung lebih tinggi termasuk ke dalam *cluster* 2. Hasil penggerombolan ini sejalan dengan hasil penelitian Safruddin, dkk (2023) yang menggunakan variabel produksi jagung dan Afdhal Nur Riyadhoh, dkk (2021) terhadap seluruh komoditas tanaman pangan di Sumatera Utara.

V. KESIMPULAN, IMPLIKASI, SARAN, DAN REKOMENDASI

Penggerombolan kabupaten/kota di Provinsi Sumatera Utara berdasarkan produktivitas dari komoditas tanaman pangan tahun 2010-2020 dilakukan menggunakan metode *average linkage* dan *K-Means Clustering* menghasilkan jumlah gerombol optimal yang sama. Metode *K-Means Clustering* lebih baik dalam menggerombolkan kabupaten/kota. *Cluster* 1 merupakan *cluster* dengan produktivitas tanaman rendah, sedangkan *cluster* 2 merupakan *cluster* dengan produktivitas tanaman pangan tinggi. Hal ini dapat dijadikan acuan bagi pemerintah dalam mengambil kebijakan untuk memaksimalkan potensi tanaman pangan kabupaten/kota di Sumatera Utara. Saran pada penelitian selanjutnya adalah dengan menambah periode data sehingga diperoleh hasil yang lebih baik.

DAFTAR PUSTAKA

- Afdhah Nur Riyadhoh, dkk. (2021). Pengelompokan Kabupaten/Kota Berdasarkan Produksi Tanaman Pangan Sumatera Utara Tahun 2020 Menggunakan Pengelompokan Hirarki Aglomeratif. KUBIK: Jurnal Publikasi Ilmiah Matematika. Vol.6 No. 2 Edisi November 2021
- Alfina, T., B. Santosa, and A. R. Barakbah (2012). Analisa perbandingan metode hierarchical clustering, k-means dan gabungan keduanya dalam cluster data (studi kasus: Problem kerja praktek teknik industri its). Jurnal Teknik ITS 1(1), A521–A525.
- BPS Provinsi Sumatera Utara. (2021). Statistik Tanaman Pangan (Padi dan Palawija) Provinsi Sumatera Utara 2021. Diakses melalui

- <https://sumut.bps.go.id/publication/2022/12/30/fb2e99417e755574546d413f/statistik-tanaman-pangan--padi-dan-palawija--provinsi-sumatera-utara-2021.html> pada 7 Januari 2023.
- Halim, N. N., & Widodo, E. (2017). Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps (SOM). *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika Dan Nilai-Nilai Islami)*, 1(1): 188–194.
- Johnson, R.A. and Winchern, D.W. 2007. “Applied Multivariate Statistical Analysis”. Pearson Education International. USA.
- Johnson, S.C. (1967). “Hierarchical Clustering Schemes”. New Jersey: Prentice-Hall, Inc.
- Mattjik, A.A., Sumertajaya I.M. (2011). *Sidik Peubah Ganda Dengan Menggunakan SAS*. Bogor: IPB Press.
- M. Khajvand dan M. J. Tarokh, (2011). “Estimating Customer Future Value of Different Customer Segments Based on Adapted RFM Model in Retail Banking Context,” *Procedia Computer Science*, vol. 3, p. 1327–1332.
- Safitri, E., dkk. (2018). Penggerombolan Provinsi di Indonesia Berdasarkan Produktivitas Tanaman Pangan Tahun 2005-2015 Menggunakan Metode K-Error. *Xplore 2* (1), 25-32.
- Safuruddin, dkk (2023). Pengelompokan Produksi Tanaman Jagung di Sumatera Utara Menggunakan Algoritma K-Medoids. *Jurnal Media Informatika Budidarma*. Volume 7, Nomor 1, Januari 2023, Page 484-491