Pengelompokan Provinsi Penghasil Padi Di Indonesia Menggunakan *Hierarchical Agglomerative Clustering*

Vigo Santri Ali*[1], Anindita Septiarini[2], Novianti Puspitasari[3]
[1,2,3]Informatika, Universitas Mulawarman, Samarinda
Jalan Kuaro, Gn. Kelua, Samarinda Ulu, Kota Samarinda, 75119
e-mail: ***[1]vigosantriali@gmail.com**, [2]anindita.septiarini@gmail.com,
[3]novia.ftik.unmul@gmail.com

Abstrak

Produksi padi merupakan salah satu indikator penting dalam ketahanan pangan di Indonesia dan memiliki karakteristik yang berbeda antar provinsi. Perbedaan tersebut perlu dianalisis untuk memperoleh pola pengelompokan provinsi yang memiliki karakteristik serupa. Penelitian ini bertujuan untuk mengelompokkan provinsi di Indonesia berdasarkan luas panen, produksi, dan produktivitas padi menggunakan metode Hierarchical Agglomerative Clustering (HAC) dengan pendekatan Average Linkage. Data yang digunakan merupakan data sekunder yang diperoleh dari instansi resmi dan dianalisis menggunakan beberapa ukuran jarak, yaitu Manhattan, Euclidean, dan Minkowski. Untuk menentukan kualitas dan jumlah cluster optimal digunakan metode Silhouette Coefficient. Hasil penelitian menunjukkan bahwa pengelompokan terbaik diperoleh pada tiga cluster, di mana setiap cluster memiliki karakteristik produksi padi yang berbeda. Cluster yang terbentuk mampu menggambarkan perbedaan tingkat produksi dan produktivitas padi antar provinsi secara jelas. Hasil pengelompokan ini diharapkan dapat menjadi bahan pertimbangan bagi pengambil kebijakan dalam perencanaan dan evaluasi sektor pertanian, khususnya komoditas padi.

***Kata Kunci**—Padi, Hierarchical Agglomerative Clustering, Average Linkage, Silhouette Coefficient*

1. PENDAHULUAN

Tanaman padi sangat penting karena menjadi sumber pangan bagi lebih dari setengah penduduk dunia [1]. Di Indonesia, nasi sebagai hasil olahan padi merupakan makanan utama sehingga memiliki peran yang sangat signifikan [2]. Dengan jumlah penduduk lebih dari 200 juta, kebutuhan padi untuk industri dan rumah tangga mencapai lebih dari 30 juta ton per tahun dan terus meningkat seiring pertumbuhan penduduk [3]. Berdasarkan hasil survei KSA, luas panen padi Indonesia tahun 2021 menurun 245,47 ribu hektar (2,30%) dibanding 2020. Jawa Timur, Jawa Tengah, dan Jawa Barat menjadi kontributor terbesar dengan luas panen masing-masing 1,75 juta, 1,70 juta, dan 1,60 juta hektar. Penurunan terbesar terjadi di Lampung, Sumatera Selatan, dan Kalimantan Selatan. Produksi padi nasional tahun 2021 sebesar 31,36 juta ton beras, turun 0,45% dari tahun sebelumnya. Sementara itu, produktivitas meningkat 1,9% dari 51,28 kuintal GKG/ha. Produktivitas tertinggi terdapat di Bali (58,83 kuintal GKG/ha) dan terendah di Kalimantan Tengah (30,28 kuintal GKG/ha) [4]. Meskipun produktivitas padi nasional mengalami peningkatan, luas panen dan produksi padi secara nasional justru mengalami penurunan masing-masing sebesar 2,30% dan 0,43% pada tahun 2021 [5]. Kondisi ini menyebabkan luas panen, produksi, dan produktivitas padi berbeda antar provinsi, sehingga diperlukan pengelompokan untuk menentukan potensi penghasil padi tertinggi [6]. Salah satu metode yang digunakan adalah *Hierarchical Agglomerative Clustering* dengan parameter jumlah *cluster*, menggunakan algoritma seperti *average linkage* [7].

Hierarchical Agglomerative Clustering (HAC) adalah metode pengelompokan yang menggabungkan objek atau *cluster* paling mirip secara bertahap hingga menjadi satu *cluster*. Proses dimulai dari setiap objek sebagai *cluster* tunggal dan digabungkan secara iteratif berdasarkan jarak tertentu. Salah satu metode yang digunakan dalam analisis ini adalah *average linkage* [8]. Penelitian lain membahas permasalahan kemiskinan sebagai salah satu tantangan sosial global dengan menerapkan metode HAC. Data yang digunakan berasal dari 34 provinsi di Indonesia dengan 7 variabel, Hasil penelitian diperoleh 3 *cluster* tingkat kemiskinan yaitu *cluster* 1 merupakan tingkat rendah dengan anggota 25 provinsi, *cluster* 2 atau tingkat sedang sebanyak 7 provinsi, dan *cluster* 3 dengan tingkat tinggi sebanyak 2 anggota [9]. Penelitian sebelumnya juga menerapkan metode HAC dalam mengelompokkan jumlah pendonor darah di wilayah Kabupaten Purwakarta. Hasil penelitian dari *cluster* data ini mendapatkan 3 *cluster*, pendonor terbanyak berada di kecamatan Purwakarta dan diikuti oleh kecamatan Bungursari. Kemudian di evaluasi menggunakan *Silhouette Coefficient* yang menghasilkan tingkat akurasi sebesar 0.7133462 [10]. Penelitian terkait pengelompokan kota/kabupaten penghasil padi di Jawa Tengah dengan menggunakan metode *K-means*. *Clustering* yang digunakan adalah metode *K-Means*. Dengan menggunakan metode ini diperoleh 3 klaster, C0 yang mencerminkan daerah dengan tingkat produktivitas padinya sedang, terdiri dari 12 daerah. C1 mencerminkan daerah yang hasil produktivitas padinya rendah terdiri dari 18 daerah. Serta C2 mencerminkan daerah yang hasil produktivitas padinya tinggi terdiri dari 12 daerah [11].

Penelitian sebelumnya menggunakan metode *K-medoids* untuk mengelompokkan provinsi berdasarkan luas panen, produksi, dan produktivitas padi. Hasilnya, *Cluster* 1 terdiri dari 30 provinsi dengan produktivitas rendah, sedangkan *Cluster* 2 terdiri dari 4 provinsi dengan produktivitas tinggi. Nilai silhouette masing-masing sebesar 0,89 (struktur sangat kuat) dan 0,66 (struktur baik) [12]. Penelitian sebelumnya juga menggunakan metode HAC untuk menganalisis dan memetakan kabupaten/kota di Provinsi Jawa Timur berdasarkan jumlah kasus penyakit. Hasilnya menunjukkan metode *average linkage* sebagai yang terbaik berdasarkan koefisien *cophenetic*, dengan jumlah *cluster* optimal sebanyak 4 berdasarkan uji validitas *connectivity*, indeks Dunn, dan *silhouette*. [13]. Penelitian sebelumnya menggunakan metode HAC dengan data jumlah penduduk kabupaten/kota di Provinsi Kalimantan Timur tahun 2005–2017. Variabel yang digunakan adalah jumlah penduduk tiap tahun, kemudian dikelompokkan menggunakan berbagai algoritma *agglomerative* seperti *single, complete, average, ward*, WPGMA, *median*, dan *centroid*. Hasil evaluasi menunjukkan nilai koefisien *cophenetic* terbaik mendekati 1 (0,99) diperoleh dari metode *autocorrelation-based distance* dengan algoritma *average linkage*. Selanjutnya, berdasarkan koefisien *silhouette*, jumlah *cluster* optimal adalah 2 [14].

Dari berbagai penelitian tersebut, penelitian menggunakan HAC untuk melakukan *clustering* provinsi penghasil padi di Indonesia dengan tujuan memberikan informasi tentang provinsi yang berpontensi sebagai lumbung padi terbesar demi menjaga ketahanan pangan nasional.

2. METODE PENELITIAN

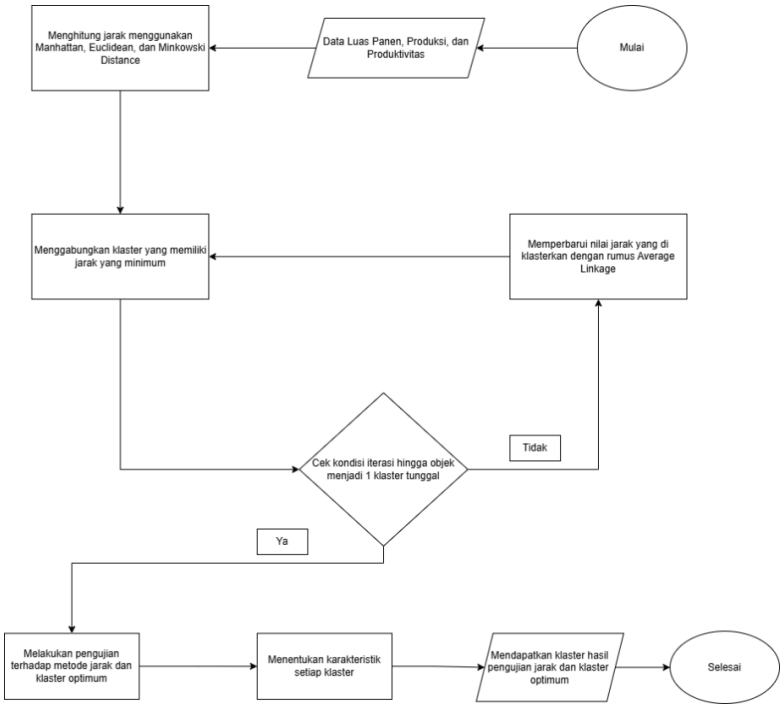
2.1 Hierarchical Agglomerative Clustering

Pada penelitian ini metode *Hierarchical Agglomerative Clustering* menggunakan *average linkage* melalui beberapa proses yang dapat dilihat pada *flowchart* Gambar 1. Alur *flowchart* Gambar 1 dijelaskan sebagai berikut:

1) Tahapan awal adalah memasukkan data luas panen rata-rata, produktivitas rata-rata, dan produksi rata-rata padi menurut Provinsi di Indonesia.
2) Kemudian, dilakukan penghitungan jarak menggunakan metode *Euclidean*, *Manhattan*, dan *Minkowski*, di mana hasil perhitungan akan menghasilkan matriks jarak untuk setiap metode tersebut.
3) Setelah memperoleh jarak dari langkah sebelumnya, langkah berikutnya adalah menerapkan metode *Hierarchical Agglomerative Clustering* dengan menggunakan *average linkage.* Pada

tahap ini, proses tersebut diulang secara berulang hingga Provinsi tergabung dalam satu klaster tunggal.

4) Melakukan tahap pengujian terhadap klaster yang terbentuk dengan menggunakan metode Koefisien *Silhouette* berdasarkan persamaan 2.7 dan menentukan karakteristik tiap klaster dengan menghitung rata-rata variabel setiap klasternya.
5) Mendapatkan hasil berupa klaster luas panen, produktivitas, dan produksi padi menurut Provinsi di Indonesia



Gambar 1. *Flowchart Hierarchical Agglomerative Clustering*

2.2 Manhattan Distance

Manhattan Distance adalah metode perhitungan jarak untuk menghitung perbedaan nilai absolut (mutlak) antara koordinat sepasang objek [15]. Rumus persamaan yang dapat digunakan untuk menghitung jarak dengan *Manhattan Distance* dapat dilihat pada persamaan (1).

$$d_{ij} = \sum_{k=1}^{d} |x_{ik} - x_{jk}| \quad (1)$$

Dengan d_{ij} adalah jarak antara objek *i* dengan *j*, x_{ik} adalah nilai objek *i* pada variabel ke-*k*, x_{jk} nilai objek *j* pada variabel ke-*k*, dan *d* adalah banyaknya data variabel.

2.3 Euclidean Distance

Euclidean distance merupakan persamaan yang menghitung jarak antar data, dengan mengukur jarak dari 2 buah titik dalam *Euclidean space* [16]. Rumus persamaan untuk menghitung *euclidean distance* dalam persamaan (2).

$$d_{ij} = \sqrt{\sum_{k=1}^{d} |x_{ik} - x_{jk}|^2} \quad (2)$$

Dengan d_{ij} adalah jarak antara objek *i* dengan *j*, x_{ik} adalah nilai objek *i* pada variabel ke-*k*, x_{jk} nilai objek *j* pada variabel ke-*k*, dan *d* adalah banyaknya data variabel.

2.4 Minkowski Distance

Minkowski distance digunakan untuk menghitung jarak dua buah atau lebih vektor[16]. Rumus persamaan untuk menghitung *minkowski distance* dalam persamaan (3).

$$d_{ij} = \left(\sum_{k=1}^{d} \left| x_{ik} - x_{jk} \right|^{p} \right)^{\left(\frac{1}{p}\right)} \quad (3)$$

Dengan d_{ij} adalah jarak antara objek *i* dengan *j*, x_{ik} adalah nilai objek *i* pada variabel ke-*k*, x_{jk} nilai objek *j* pada variabel ke-*k*, *d* adalah banyaknya data variabel, dan *p* parameter order, jika bernilai 1 maka perhitungan sama dengan jarak *Manhattan* dan ketika bernilai 2 maka sama dengan jarak *Euclidean.*

2.5 Average Linkage

Metode *average linkage* menggunakan Prinsip ukuran jarak rata-rata antar tiap pasangan objek yang mungkin. Pengelompokan dimulai dengan mencari nilai minimum dari $\boldsymbol{d_{(AB)}}$ dan menggabungkan objek-objek yang saling berdekatan, misal dan untuk mendapatkan *Cluster* [17]. Rumus persamaan untuk menghitung *average linkage* dalam persamaan (4).

$$d_{(AB)C} = \frac{d_{(AC)} + d_{(BC)}}{n_{(AB)} n_C} \quad (4)$$

Dengan $n_{(AB)}$ adalah banyaknya anggota dalam klaster (AB) dan n_c adalah banyaknya anggota dalam klaster C

2.6 Silhouette Coefficient

Silhouette Coefficient merupakan salah metode yang digunakan untuk menentukan banyak *cluster* yang Optimal berdasarkan validasi *cluster*, seberapa banyak *cluster* yang optimal [12]. Rumus persamaan untuk menghitung *Silhouette Coefficient* dalam persamaan (5).

$$s_i = \frac{(b_i - a_i)}{\max(a_i b_i)} \quad (5)$$

Dengan a_i adalah rata-rata ke objek ke-*i* pada kelompok yang sama, b_i adalah rata-rata ke objek ke-*i* pada kelompok yang berbeda, dan s_i adalah nilai *Silhouette Coefficient.*

2.7 Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data luas panen, produksi, dan produktivitas padi di Indonesia pada tahun 2021-2023 yang di dapat melalui *website* resmi Badan Pusat Statistik Indonesia. Penjelasan data berikut dapat dilihat pada Tabel 1.

Tabel 1. Data Penelitian

Provinsi	Luas panen (ha)			Produktivitas (ku/ha)			Produksi (ton)		
	2021	2022	2023	2021	2022	2023	2021	2022	2023
Aceh	297058,38	271750,2	254318,63	55,03	55,55	54,79	1634639,6	1509456	1393474,11
Sumatera Utara	385405	411462,1	404472,52	52	50,76	51,44	2004142,51	2088584	2080663,46
Sumatera Barat	272391,95	271883,1	296492,13	48,36	50,52	49,16	1317209,38	1373532	1457502,44
Riau	53062, 35	51054,04	5182 0,64	40,98	41,83	40,37	217458,87	213557,2	209190,02

Jambi	64412,26	60539,59	61378,11	46,29	45,88	44,73	298149,25	277743,8	274557,09
Sumatera Selatan	496241,65	513378,2	502162,22	51,44	54,06	55	2552443,19	2775069	2762059,57
Bengkulu	55704,69	57151,84	56803,3	48,67	49,27	48,82	271117,19	281610,1	277310,01
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Gorontalo	48713,5	46823,47	48829,99	48,12	51,29	49,8	234392,86	240134,5	243193,49
Sulawesi Barat	59763,18	69323,95	59116,39	52,05	50,99	49,74	311072,46	353513,3	294026,68
Maluku	28319,75	23987,82	22615,89	41,24	38,6	36,73	116803,67	92601,06	83065,17
Maluku Utara	7781,96	6416,45	7684,39	36,05	38,16	36,66	28050,8	24486,03	28168,81
Papua Barat	6414,94	5460,59	5532,26	41,98	43,89	43,04	26926,93	23963,92	23808,11
Papua	64984,9	49741,91	49322,77	44,05	38,99	40,57	286279,8	193943,5	200115,34
Indonesia	10411801,22	10452672	10196886,77	52,26	52,38	52,59	54415294,22	5474897	53625539,51

Berdasarkan Tabel 1 dapat dilihat bahwa pengelompokan provinsi penghasil padi di Indonesia menggunakan *Hierarchical Agglomerative Clustering*, data terdiri dari 34 provinsi dengan luas panen, produktivitas, dan produksi padi di Indonesia dari tahun 2021 sampai tahun 2023.

2.8 Perancangan Data

Data penelitian berupa luas panen, produktivitas, dan produksi padi pada 34 provinsi di Indonesia selama periode 2021–2023 yang diperoleh dari Badan Pusat Statistik. Untuk menyederhanakan analisis, data tiap variabel dihitung nilai rata-rata per provinsi selama tiga tahun terakhir (2021–2023). Data yang telah dikumpulkan akan dilakukan beberapa proses sebelum dilakukan proses *clustering.* Proses-proses yang dilakukan yaitu:

a. Data *selection*
   Dari data luas panen, produksi, dan produktivitas padi tahun 2021 sampai dengan 2023 kemudian di *selection* menjadi beberapa atribut berdasarkan kebutuhan pada penelitian. Variabel yang digunakan adalah X1(Luas Panen Rata-rata), X2 (Produktivitas Rata-rata), X3(Produksi Rata-rata). Provinsi yang ada di Indonesia diubah menjadi kode yang akan digunakan adalah P1 (Aceh), P2 (Sumatra Utara), P3 (Sumatra barat), P4 (Riau), P5 (Jambi), P6 (Sumatra Selatan), P7 (Bengkulu), P8 (Lampung), P9 (Kep.Bangka Belitung), P10 (Kep.Riau), P11 (DKI Jakarta), P12 (Jawa Barat), P13 (Jawa Tengah), P14 (DI Yogyakarta), P15 (Jawa Timur), P16 (Banten), P17 (Bali), P18 (Nusa Tenggara Barat), P19 (Nusa Tenggara Timur), P20 (Kalimantan barat), P21 (Kalimantan Tengah), P22 (Kalimantan Selatan), P23 (Kalimantan Timur), P24 (Kalimantan Utara), P25 (Sulawesi Utara), P26 (Sulawesi Tengah), P27 (Sulawesi Selatan), P28 (Sulawesi Tenggara), P29 (Gorontalo), P30 (Sulawesi Barat), P31 (Maluku), P32 (Maluku Utara), P33 (Papua Barat), P34 (Papua), dan P35 (Indonesia).
b. Data *integration*
   Data yang telah diseleksi dari *Website* Badan Pusat Statistik di integrasikan dalam satu tabel untuk digunakan dalam proses pengelompokan Provinsi penghasil padi di Indonesia. Tabel terdiri atas 4 kolom yaitu Provinsi di Indonesia sebagai kolom pertama, Luas Panen rata-rata sebagai kolom kedua, Produktivitas rata-rata sebagai kolom ketiga, dan Produksi rata-rata sebagai kolom keempat.
c. Data *cleaning*
   Proses *data cleaning* dilakukan pengecekan data secara manual dengan membuang data yang tidak relevan atau tidak diperlukan dalam proses penelitian. Dengan mecari data yang tidak

relevan dengan tujuan penelitian, menghasilkan bahwa data ”Indonesia” bukan termasuk dalam bagian Provinsi di Indonesia sehingga dilakukannya proses *cleaning*

d. Normalisasi Data

Normalisasi data dilakukan untuk mencegah adanya perbedaan yang jauh antar variabel, normalisasi akan mengubah nilai menjadi 0 sampai dengan 1. Terdapat berbagai metode normalisasi data, seperti *Min-max Normalization, Z-score Normalization,* dan *Decimal scaling* [18]. Pada tahap ini akan dilakukan proses normalisasi *min-max* pada data yang telah dikumpulkan dengan menggunakan rumus persamaan (6).

$$nilai\ normalisasi = \frac{(x_{awal} - x_{min})}{(x_{max} - x_{min})} \quad (6)$$

Keterangan:

x_{awal} : nilai awal
x_{min} : nilai terkecil
x_{max} : nilai terbesar

Normalisasi *min-max* dilakukan untuk meratakan rentang data diantara 3 variabel agar mempermudah perhitungan.

3. HASIL DAN PEMBAHASAN

Pada bagian ini disajikan hasil penelitian. Tahap awal dilakukan penggabungan data (data *integration*) ke dalam satu tabel berdasarkan luas panen rata-rata, produksi rata-rata, dan produktivitas rata-rata padi tahun 2021–2023 untuk proses pengelompokan provinsi. Selanjutnya dilakukan data *cleaning* dengan menghapus P35 (Indonesia) yang tidak relevan dengan penelitian. Hasil kedua tahapan dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pengolahan Data

Provinsi	X1	X2	X3
P1	274.375,74	55,12	1.512.523,24
P2	400.446,54	51,40	2.057.796,66
P3	280.255,73	49,35	1.382.747,94
P4	51.979,01	41,06	213.402,03
P5	62.109,99	45,63	283.483,38
P6	503.927,36	53,50	2.696.523,92
P7	56.553,28	48,92	276.679,10
P8	513.533,42	51,29	2.634.131,13
P9	16.264,26	40,58	65.807,39
P10	196,35	30,08	595,01
⋮	⋮	⋮	⋮
P27	998.784,71	51,37	5.131.300,86
P28	120.637,82	41,20	497.119,38
P29	48.122,32	49,74	239.240,28
P30	62.734,51	50,93	319.537,48
P31	24.974,49	38,86	97.489,97
P32	7.294,27	36,96	26.901,88
P33	5.802,60	42,97	24.899,65
P34	54.683,19	41,20	226.779,55

Langkah selanjutnya adalah menormalisasikan data hasil dari data *cleaning* agar mencegah adanya perbedaan yang jauh antar variabel, normalisasi akan mengubah nilai menjadi 0 sampai dengan 1 menggunakan persamaan (6). Hasil dari normalisasi data dapat dilihat pada Tabel 3.

Tabel 3. Hasil Normalisasi Data

Provinsi	X1	X2	X3
P1	0,16047	0,83860	0,15692
P2	0,23426	0,71392	0,21351

P3	0,16392	0,64516	0,14345
P4	0,03031	0,36767	0,02209
P5	0,03624	0,52082	0,02936
P6	0,29483	0,78424	0,27980
⋮	⋮	⋮	⋮
P29	0,02805	0,65822	0,02477
P30	0,03660	0,69807	0,03310
P31	0,01450	0,29389	0,01006
P32	0,00415	0,23027	0,00273
P33	0,00328	0,43163	0,00252
P34	0,03189	0,37247	0,02347

Berdasarkan tahapan penelitian yang dilakukan didapatkan hasil *cluster* untuk skenario 2 *cluster*, 3 *cluster* dan 4 *cluster* dengan 3 metode jarak yaitu *Manhattan*, *Euclidean*, dan *Minkowski*. Hasil pengelompokan 3 *cluster* dapat dilihat pada Tabel 4.

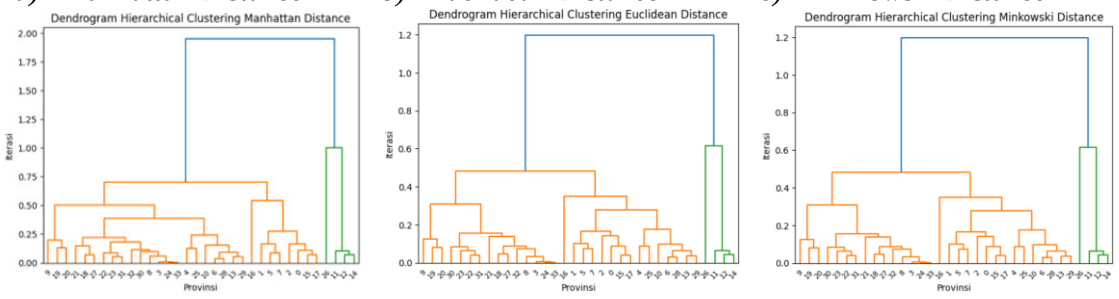
Tabel 4. Hasil Pengelompokan *3 Cluster*

No.	Metode Jarak	*Cluster* 1	*Cluster* 2	*Cluster* 3
1.	*Manhattan Distance*	P1,P2,P3,P4,P5,P6,P7,P8, P9,P10,P11,P14,P16,P17, P18,P19,P20,P21,P22,P23 ,P24,P25,P26,P28,P29,P3 0,P31,P32,P33,P34	P12, P13, P15	P27
2.	*Euclidean Distance*	P12,P13,P15,P27	P1,P2,P3,P5,P6,P7,P8, P11,P14,P16,P17,P18,P 26,P29,P30	P4,P9,P10,P19,P20,P21,P 22,P23,P24,P25,P28,P31, P32,P33,P34
3.	*Minkowski Distance*	P12,P13,P15,P27	P1,P2,P3,P5,P6,P7,P8, P11,P14,P16,P17,P18,P 26,P29,P30	P4,P9,P10,P19,P20,P21,P 22,P23,P24,P25,P28,P31, P32,P33,P34

Berdasarkan Tabel 4 dapat dilihat bahwa hasil 3 *cluster* pada 3 metode jarak *Manhattan* memiliki *cluster* 3 yaitu P27, *Euclidean* dan *Minkowski* memiliki *cluster* 3 yang sama yaitu P4, P9, P10, P19, P20, P21, P22, P23, P24, P25, P28, P31, P32, P33, P34.

Selanjutnya, Gambar 2 menampilkan dendrogram hasil penerapan *Hierarchical Agglomerative Clustering* menggunakan tiga pengukuran jarak yang menggambarkan proses penggabungan antar provinsi berdasarkan tingkat kemiripan karakteristik luas panen, produksi, dan produktivitas padi. Pada dendrogram tersebut, setiap provinsi awalnya berdiri sebagai satu *cluster* tersendiri, kemudian secara bertahap digabungkan dengan provinsi lain yang memiliki jarak paling kecil. Semakin rendah ketinggian penggabungan pada dendrogram menunjukkan semakin tinggi tingkat kemiripan antar provinsi, sedangkan penggabungan pada ketinggian yang lebih tinggi menunjukkan perbedaan karakteristik yang semakin besar.

a) Manhattan Distance *b) Euclidean Distance* *c) Minkowski Distance*



Gambar 2. Tampilan Dendogram Hasil 3 Pengukuran Jarak

Pengujian untuk mendapatkan hasil *cluster* terbaik dilakukan dengan *Silhouette coefisient*. Nilai *Silhouette Coefisient* dapat dilihat pada Tabel 5.

Tabel 5. Hasil Nilai *Silhouette Coefficient*

No.	Metode Jarak	2 *Cluster*	3 *Cluster*	4 *Cluster*
1	*Manhattan*	0,70016	**0,82784**	0,72297
2	*Euclidean*	0,68220	0,57273	0,74462
3	*Minkowski*	0,66914	0,58580	0,75131

Hasil perhitungan yang telah dilakukan pada Tabel 7 dapat diketahui bahwa perhitungan dengan menggunakan metode jarak *Manhattan* dengan jumlah 3 *cluster* dengan nilai *Silhouette Coefficient* 0,82784 di mana nilai ini merupakan nilai yang paling mendekati nilai 1 dibandingkan dengan metode jarak yang berbeda dan *cluster* yang berbeda sehingga dapat dihasilkan *cluster* optimal tiap provinsi yang dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pengelompokan Jarak *Manhattan* 3 *Cluster*

Cluster	Provinsi
Cluster 1	P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P14, P16, P17, P18, P19, P20, P21, P22, P23, P24, P25, P26, P28, P29, P30, P31, P32, P33, P34
Cluster 2	P12, P13, P15
Cluster 3	P27

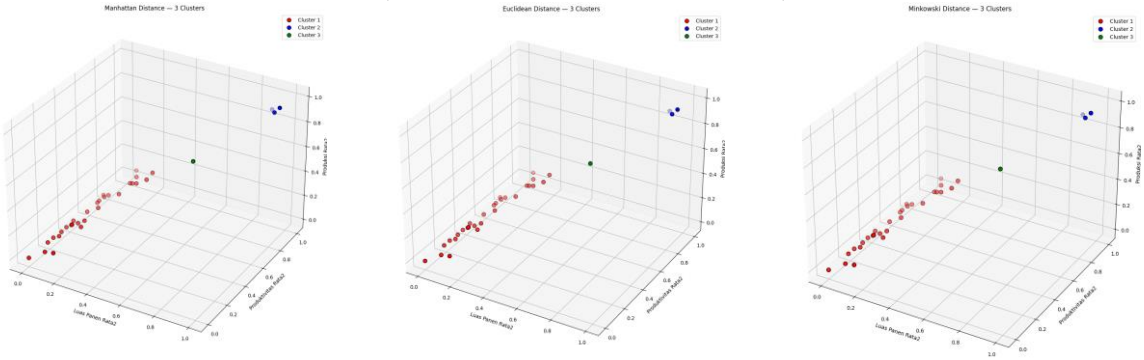
Pada Tabel 6 dapat dilihat bahwa *cluster* 1 memiliki 30 provinsi, lalu pada *cluster* 2 memiliki 3 provinsi yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur dan pada *cluster* 3 terdapat Provinsi Sulawesi Selatan. Hasil *cluster* terbaik kemudian diinterpretasi dengan menghitung rata-rata setiap variabel pada setiap *cluster*. Nilai *centroid* pada setiap *cluster* dapat dilihat pada Tabel 7.

Tabel 7. Nilai *Centroid* 3 *Cluster* Metode *Average Linkage*

Centroid	V1	V2	V3
Nilai *Centroid Cluster* 1	0,08485	0,48886	0,07237
Nilai *Centroid Cluster* 2	0,97532	0,88142	0,97538
Nilai *Centroid Cluster* 3	0,58446	0,71291	0,53249

Berdasarkan Tabel 7, nilai *centroid* menunjukkan bahwa *Cluster* 1 memiliki nilai terendah pada seluruh variabel, sehingga merupakan kelompok provinsi penghasil padi terendah (30 provinsi). *Cluster* 2 memiliki nilai tertinggi pada seluruh variabel, sehingga menjadi kelompok penghasil padi tertinggi (3 provinsi). Sementara itu, *Cluster* 3 memiliki nilai di antara keduanya, yang mengindikasikan potensi menjadi provinsi penghasil padi tinggi. Visualisasi *clustering* dengan tiga pengukuran jarak dan 3 *cluster* ditampilkan dalam *scatter plot* 3D pada Gambar 4 berdasarkan luas panen, produktivitas, dan produksi rata-rata padi.

a) *Manhattan* 3 *cluster* b) *Euclidean* 3 *cluster* c) *Minkowski* 3 *cluster*



Gambar 3. *Plot* 3 *Cluster*

Dari Gambar 4, terlihat bahwa setiap titik merepresentasikan provinsi, sedangkan warna menunjukkan keanggotaan *cluster*. *Cluster* 1 (merah) berisi provinsi dengan nilai rendah hingga sedang yang mengelompok di bagian bawah grafik. *Cluster* 2 (biru) terdiri dari provinsi dengan nilai sangat tinggi sehingga terpisah jauh sebagai daerah unggulan. Sementara itu, *Cluster* 3 (hijau) berada di antara keduanya dan merepresentasikan provinsi dengan karakteristik menengah atau berpotensi berkembang.

4. KESIMPULAN

Hasil pengujian menggunakan *Silhoutte Coefficient* terhadap akurasi 3 metode jarak yaitu *Manhattan*, *Euclidean*, dan *Minkowski* dengan menggunakan metode *Hierarchical Agglomerative Clustering* dengan *Average linkage* yang menghasilkan *cluster* yang paling optimal dengan jarak Manhattan 3 *cluster*. Hasil *cluster* 1 merupakan provinsi penghasil padi terendah beranggotakan 30 provinsi. *Cluster* 2 merupakan provinsi penghasil padi tertinggi yang terdiri dari 3 provinsi yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur. *Cluster* 3 merupakan provinsi yang memiliki potensi untuk berkembang yaitu Sulawesi Selatan.

5. SARAN

Pada penelitian selanjutnya diharapkan dapat menggunakan metode lain seperti metode *Ward*, *Single Linkage* ataupun *Complete Linkage* dan menggunakan metode jarak yang berbeda dari metode jarak yang telah digunakan seperti *Measure Distance* serta menguji jumlah *cluster* yang lebih optimal dengan *Partition Coeffiecient Index* (PCI).

DAFTAR PUSTAKA

[1] M. A. Ningrat, C. Diana, dan Y. Y. Makabori, "Pertumbuhan dan Hasil Tanaman Padi (Oryza sativa L.) pada Berbagai Sistem Tanam di Kampung Desay, Distrik Prafi, Kabupaten Manokwari," *Prosiding Seminar Nasional Pembangunan dan Pendidikan Vokasi Pertanian*, vol. 2, no. 1, hlm. 325–332, Sep 2021, doi: 10.47687/snppvp.v2i1.191.

[2] D. Septiadi dan U. Joka, "Analisis Respon dan Faktor-Faktor yang Mempengaruhi Permintaan Beras Indonesia," *AGRIMOR*, vol. 4, no. 3, hlm. 42–44, Jul 2019, doi: 10.32938/ag.v4i3.843.

[3] A. Poernomo dan H. Winarto, "Kemampuan Produksi Sumber Pangan Pokok Dan Non Biji-Bijian Terhadap Ketahanan Pangan Kabupaten Banyumas," *Majalah Ilmiah Manajemen dan Bisnis*, vol. 17, hlm. 1–12, 2020.

[4] Badan Pusat Statistik, "Luas Panen dan Produksi Padi di Indonesia 2021."

[5] V. budi Kusnandar, "Produktivitas Padi Indonesia Meningkat 1,9% pada 2021," https://databoks.katadata.co.id/datapublish/2022/03/02/produktivitas-padi-indonesia-meningkat-19-pada-2021.

[6] S. Ani Ritonga, M. Safii, I. Parlina, H. Satria Tambunan, P. Studi Sistem Informasi, dan S. A. Tunas Bangsa Pematangsiantar Jln Jendral Sudirman Blok No, "Prosiding Seminar Nasional Riset Information Science (SENARIS) Teknik Data Mining dalam Mengelompokkan Produktivitas Padi Menurut Provinsi Menggunakan K-Medoids," 2019, [Daring]. Tersedia pada: https://www.bps.go.id.

[7] I. M. Sabara, "Analisis Agglomerative Hierarchical Clustering Berdasarkan Pengurutan Parsial Graff Hasse Terhadap Indikator Kemiskinan di Jawa Timur," 2022.

[8] S. Fikri Mu'afa, N. Ulinnuha, I. Sunan, dan A. Surabaya, "Perbandingan Metode Single Linkage, Complete Linkage Dan Average Linkage dalam Pengelompokan Kecamatan Berdasarkan Variabel Jenis Ternak Kabupaten Sidoarjo," vol. 4, no. 2, 2019.

[9] E. Widodo, P. Ermayani, L. N. Laila, dan A. T. Madani, "Pengelompokkan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative Clustering (Indonesian Province Grouping Based on Poverty Level Using Hierarchical Agglomerative Clustering Analysis)," 2021.

[10] D. Iyan Yulianti, T. Iman Hermanto, dan M. Defriani, "Rekayasa Teknik Informatika dan Informasi Analisis Clustering Donor Darah dengan Metode Agglomerative Hierarchical Clustering," *Media Online)*, vol. 3, no. 6, hlm. 441, 2023, [Daring]. Tersedia pada: https://djournals.com/resolusi

[11] S. Wijayanto, M. Yoka Fathoni, J. DI Panjaitan No, K. Purwokerto Selatan, K. Banyumas, dan J. Tengah, "Pengelompokkan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means," 2021.

[12] J. Terapan, S. & Teknologi, E. Syafaqoh, N. Sunan, dan A. Surabaya, "Pengelompokan Provinsi Di Indonesia Berdasarkan Luas Panen, Produksi, Dan Produktivitas Padi Menggunakan Algoritma K-Medoid," *Fakultas Sains dan Teknologi-Universitas PGRI Kanjuruhan Malang*, vol. 5, no. 3, hlm. 2023, 2023.

[13] A. Septianingsih, S. A. Pertama, D. Kependudukan, P. Sipil, dan K. Tangerang, "Pemetaan Kabupaten Kota di Provinsi Jawa Timur Berdasarkan Tingkat Kasus Penyakit Menggunakan Pendekatan Agglomerative Hierarchical Clustering," vol. 3, no. 2, 2022, doi: 10.46306/lb.v3i2.

[14] A. Tri, R. Dani, S. Wahyuningsih, dan N. A. Rizki, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura Journal of Mathematics*, vol. 1, 2019, [Daring]. Tersedia pada: http://ejurnal.ung.ac.id/index.php/jjom,P-

[15] D. Ramadhaniaty, N. Puspitasari, dan A. Septiarini, "Klasterisasi Status Keberlanjutan Karyawan Kontrak Menggunakan K-Medoids Clustering Contract Employees' Sustainability Status Using K-Medoids," *TELKA*, vol. 10, no. 2, hlm. 97–108, 2024.

[16] Vicky Pranandika Wijaksana, "Penerapan Metode K-Means Clustering Status Gizi Balita Di UPT Puskesmas Barong Tongkok," *Data Sciences Indonesia (DSI)*, vol. 5, no. 1, hlm. 156–165, Jul 2025, doi: 10.47709/dsi.v5i1.6517.

[17] A. F. Dewi dan K. Ahadiyah, "Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan," *Zeta - Math Journal*, vol. 7, no. 2, hlm. 57–63, Nov 2022, doi: 10.31102/zeta.2022.7.2.57-63.

[18] Suryanegara, Adiwijaya, dan Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, hlm. 114–122, Feb 2021, doi: 10.29207/resti.v5i1.2880.