

SKRIPSI
ANALISIS *CLUSTERING* MENGGUNAKAN *AGGLOMERATIVE*
***HIERARCHICAL CLUSTER* PADA DATA HASIL PRODUKSI**
PERTANIAN DI SULAWESI BARAT TAHUN 2023



SATRIWANTI
E0221567

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SULAWESI BARAT

2025

SKRIPSI
ANALISIS *CLUSTERING* MENGGUNAKAN *AGGLOMERATIVE*
***HIERARCHICAL CLUSTER* PADA DATA HASIL PRODUKSI**
PERTANIAN DI SULAWESI BARAT TAHUN 2023



Diajukan sebagai salah satu syarat memperoleh gelar Sarjana pada Program Studi
Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sulawesi Barat

SATRIWANTI

E0221567

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SULAWESI BARAT
2025

SURAT PERNYATAAN

Yang bertanda tangan dibawa ini:

Nama : Satriwanti
Tempat/Tgl Lahir : Peleboan, 01 Juli 2003
NIM : E0221567
Program Studi : Statistika

Menyatakan bahwa tugas akhir dengan judul “Analisis *Clustering* Menggunakan *Agglomerative Hierarchical Cluster* Pada Data Hasil Produksi Pertanian di Sulawesi Barat Tahun 2023” disusun berdasarkan prosedur ilmiah yang telah melalui pembimbingan dan bukan merupakan plagiat dari karya ilmiah/naskah yang lain. Apabilah di kemudian hari terbukti bahwa pernyataan ini tidak benar, maka saya bersedia menerima sanksi sesuai yang berlaku.

Majene, Selasa 5 Agustus 2025

yang menyatakan,



Satriwanti

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Satriwanti
NIM : E0221567
Judul : *Analisis Clustering Menggunakan Agglomerative Hierarchical Cluster Pada Data Hasil Produksi Pertanian di Provinsi Sulawesi Barat Tahun 2023.*

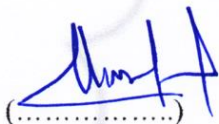
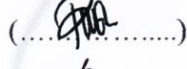
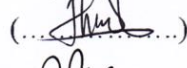
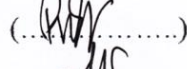
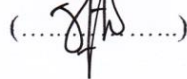
Telah berhasil dipertanggungjawabkan dihadapan Tim Penguji (Nomor SK : 90/UN55.7/HK.04/2023, tanggal 7 November 2024) dan diterima sebagai bagian persyaratan memperoleh gelar Sarjana S.Stat pada Pogram Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sulawesi Barat.

Disahkan Oleh:
Dekan FMIPA
Universitas Sulawesi Barat



Musafira, S.Si., M.Sc.
NIP. 19770911200660422002

Tim Penguji:

Ketua Penguji	: Musafira, S.Si., M.Sc.	()
Sekretaris	: Muh. Hijrah, S.Pd., M.Si.	()
Pembimbing 1	: Putri Indi Rahayu, S.Si., M.Stat.	()
Pembimbing 2	: Retno Mayapada, S.Si., M.Si.	()
Penguji 1	: Muh. Hijrah, S.Pd., M.Si.	()
Penguji 2	: Reski Wahyu Yanti, S.Si., M.Si.	()
Penguji 3	: Muhammad Hidayatullah, S.Pd., M.Kom.	()

ABSTRAK

Sektor pertanian di Indonesia masih menjadi sektor penting bagi perekonomian bangsa dan negara, dimana Indonesia dikenal sebagai negara agraris dengan berbagai produksi pertaniannya. Berdasarkan Badan Pusat Statistik (BPS) Provinsi Sulawesi Barat pada publikasi Potensi Pertanian tahun 2023, tercatat bahwa masyarakat masih menjadikan pertanian sebagai tumpuan utama ekonomi. Tujuan dari penelitian ini adalah untuk mengelompokkan kecamatan di Provinsi Sulawesi Barat berdasarkan hasil produksi pertanian tahun 2023. Penelitian ini menerapkan metode *Agglomerative Hierarchical Cluster* pada pengelompokan kecamatan. Variabel sebanyak 69 kecamatan di Provinsi Sulawesi Barat di peroleh hasil nilai korelasi *Chopenetic* yaitu metode *Average Linkage* sebesar 0,9658 yang paling mendekatii 1, dan diperoleh nilai indeks *Connectivity*, *Dunn*, *Silhouette*, WSS, yaitu 2 *cluster* sehingga dipilih sebagai metode pengelompokan terbaik, diantaranya *cluster* 1 pada komoditas variabel durian, kopi, kakao, terung, jahe, bayam, dan rambutan memiliki karakteristik tinggi dibandingkan dengan komoditas tanaman yang serupa di *cluster* 2. Sedangkan pada *cluster* 2, hanya terdiri dari Kecamatan Banggae Timur, menunjukkan hasil produksi pertanian yang menonjol pada variabel ubi kayu, kelapa, cabai besar, cabai rawit, lengkuas, mangga, pisang dan pepaya, dibandingkan kecamatan-kecamatan lain di Provinsi Sulawesi Barat.

Kata Kunci: *Agglomerative Hierarchical Clustering*, *Average Linkage*, Hasil Produksi Pertanian.

ABSTRACT

The agricultural sector in Indonesia is still an important sector for the nation's economy, Indonesia is known as an agrarian country with a variety of agricultural production. Based on the Central Bureau of Statistics (BPS) West Sulawesi Province in the publication of Agricultural Potential in 2023, it was noted that the community still made agriculture the main foundation of the economy. The purpose of this research is to group sub-districts in West Sulawesi Province based on agricultural production in 2023. This research applies the agglomerative hierarchical cluster method to grouping sub-districts. Variables as many as 69 sub-districts in West Sulawesi Province obtained the results of the Chopenetic correlation value, namely the Average Linkage method of 0.9658 which is the most climbing 1, and obtained the index value of Connectivity, Dunn, Silhouette, WSS, which is 2 clusters so that it is chosen as the best clustering method, including cluster 1 on variable commodities durian, coffee, cocoa, eggplant, ginger, spinach, and rambutan have high characteristics compared to similar plant commodities in cluster 2. Meanwhile, cluster 2, which only consists of East Banggae sub-district, shows prominent agricultural production results in the variables of cassava, coconut, large chili, cayenne pepper, galangal, mango, banana and papaya, compared to other sub-districts in West Sulawesi Province.

Keywords: *Agglomerative Hierarchical Clustering, Average Linkage, Agricultural Production.*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sektor pertanian di Indonesia masih menjadi sektor penting bagi perekonomian bangsa dan negara, dimana Indonesia dikenal sebagai negara agraris dengan berbagai produksi pertanian. Pertanian dapat didefinisikan sebagai tindakan manusia yang memanfaatkan sumber daya hayati untuk menghasilkan bahan pangan, bahan baku industri, energi, dan juga untuk mengelola lingkungan hidupnya (Tendean & Purba, 2020). Selain itu, pertanian juga merupakan salah satu sektor pendapatan masyarakat yang memiliki peranan penting di Indonesia dikarenakan masyarakat mayoritas sebagai petani (Mukuan *et al.*, 2022).

Berdasarkan Badan Pusat Statistik (BPS), Provinsi Sulawesi Barat pada publikasi Potensi Pertanian tahun 2023, tercatat bahwa masyarakat masih menjadikan pertanian sebagai tumpuan utama ekonomi. Hal ini dikarenakan kondisi alam dan keadaan geografis yang mendukung aktivitas pertanian, antara lain tanah yang subur, iklim yang relatif stabil, serta ketersediaan sumber daya air yang melimpah, yang mampu menjadikan wilayah Sulawesi Barat sangat strategis untuk berbagai jenis komoditas pertanian. Pada tahun 2023, sektor pertanian berkontribusi sekitar 44,72% terhadap Produk Domestik Regional Bruto (PDRB) Sulawesi Barat yang menunjukkan bahwa sektor ini masih menjadi tulang punggung ekonomi daerah.

Sektor pertanian merupakan salah satu indikator yang mempengaruhi Indeks Ketahanan Pangan, dikarenakan hasil pangan lebih banyak diproduksi dari hasil produksi pertanian (Andaresta, 2024). Berdasarkan data dari Badan Pangan Nasional tahun 2023, skor Indeks Ketahanan Pangan Provinsi Sulawesi Barat secara nasional masih rendah yaitu urutan 21 dari 38 provinsi dengan skor 73,03%.

Salah satu langkah yang bisa dilakukan untuk meningkatkan ketahanan pangan adalah dengan mengetahui potensi komoditas pertanian yang dapat dilakukan dengan mengelompokkan wilayah yang memiliki potensi hampir sama kedalam beberapa *cluster* (Nur *et al.*, 2021). Oleh karena itu, untuk mengetahui

potensi yang lebih dominan dan dapat memaksimalkan potensi komoditas pertanian di Sulawesi Barat, dapat dilakukan analisis *cluster*.

Analisis *cluster* didefinisikan sebagai metode analisis statistika yang didasarkan pada bagaimana masing-masing objek memiliki karakteristik yang mirip dan dapat digunakan untuk menempatkan kumpulan objek ke dalam dua grup bahkan lebih (Ulinnuha & Veriani, 2020). Analisis ini dapat dilakukan dengan dua metode yaitu hierarki dan non-hierarki. Metode hierarki adalah analisis yang mengelompokkan dengan cara mengukur jarak kedekatan pada setiap objek yang kemudian membentuk sebuah dendrogram. Sementara itu, metode non-hierarki digunakan untuk mengelompokkan objek, dimana jumlah *cluster* yang akan dibentuk sudah ditentukan sebelumnya. Contoh metode non-hierarki yaitu *k-means*, *k-median* dan *fuzzy*. Kelebihan metode hierarki dibandingkan dengan metode non-hierarki adalah proses pengolahan data yang singkat dan hasil yang berbentuk berupa tingkatan sehingga menghemat waktu (Faisal & Rifai, 2023).

Pengelompokan *hierarchical* secara umum terbagi menjadi dua jenis yaitu *agglomerative* dan *divisive*. Namun, metode *agglomerative* lebih umum digunakan dibandingkan metode *divisive* (Kaufman & Rousseeuw, 2009). Beberapa contoh metode *Agglomerative Hierarki* diantaranya adalah metode *Single Linkage*, metode *Complete Linkage*, metode *Average Linkage*, metode *Centroid*, metode *Ward*, dan metode *Median Linkage* (Widyadhan *et al.*, 2021).

Beberapa penelitian telah menggunakan analisis *clustering* seperti yang dilakukan oleh Oktavianty *et al.*, (2019), yakni pengelompokan kabupaten/kota di Sulawesi berdasarkan indikator pendidikan menggunakan analisis *cluster Average Linkage* dan *Median Linkage*. Hasil penelitiannya menunjukkan bahwa metode terbaik yaitu *Median Linkage*. Penelitian selanjutnya dilakukan oleh Ulinnuha dan Veriani (2020), yaitu analisis *cluster* dalam pengelompokan provinsi di Indonesia berdasarkan variabel penyakit menular menggunakan metode *Complete Linkage*, *Average Linkage* dan *Ward*. Hasil penelitian menunjukkan metode terbaik yaitu metode *Ward*. Penelitian lainnya yaitu Latif (2024), mengelompokkan puskesmas di Kabupaten Polewali Mandar Tahun 2023 berdasarkan indikator gizi buruk balita

dengan Model *Agglomerative Hierarchical Clustering*, dan diperoleh metode terbaik yaitu metode *Average Linkage*.

Berdasarkan uraian diatas, maka penulis tertarik melakukan pengelompokan kecamatan yang ada di Provinsi Sulawesi Barat berdasarkan komoditas hasil produksi pertanian seperti durian, ubi kayu, kelapa, kopi, kakao, cabai besar, cabai rawit, terung, jahe, lengkuas, mangga, pisang, pepaya, dan rambutan menggunakan metode *Average Linkage*, *Ward* dan *Median Linkage*. Oleh karena itu, penulis mengangkat sebuah judul yaitu Analisis *Clustering* menggunakan *Agglomerative Hierarchical Cluster* pada Data Hasil Produksi Pertanian di Sulawesi Barat Tahun 2023.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang diatas, maka diambil rumusan masalah dalam penelitian ini adalah :

1. Metode manakah yang paling efektif di antara *Average Linkage*, *Median Linkage*, dan *Ward* dalam pengelompokan kecamatan di Provinsi Sulawesi Barat berdasarkan hasil produksi pertanian tahun 2023?
2. Bagaimana karakteristik pengelompokan kecamatan di Provinsi Sulawesi Barat menggunakan metode terbaik?

1.3 Tujuan Penelitian

Berikut ini adalah tujuan yang ingin dicapai dalam penelitian ini yaitu:

1. Untuk mengetahui metode yang paling efektif di antara *Average Linkage*, *Median Linkage*, dan *Ward* dalam pengelompokan kecamatan di Provinsi Sulawesi Barat berdasarkan hasil produksi pertanian tahun 2023.
2. Untuk mengetahui karakteristik pengelompokan kecamatan di Provinsi Sulawesi Barat menggunakan metode terbaik.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Bagi penulis, dapat memberikan kontribusi ilmu pengetahuan dalam bidang statistika terkait analisis *clustering* dengan metode *Average Linkage*, *Median Linkage* dan metode *Ward*.
2. Bagi instansi, dapat menjadi bahan referensi dan pertimbangan bagi penelitian selanjutnya terkait analisis *clustering*.

1.5 Batasan Masalah

Untuk mencapai penelitian yang diharapkan maka perlu melakukan pembatasan permasalahan pada penelitian ini sebagai berikut:

1. Studi kasus yang digunakan dalam penelitian ini dibatasi pada 69 kecamatan dan 15 jenis hasil produksi di Provinsi Sulawesi Barat.
2. Data yang digunakan adalah data sekunder dari *website* Badan Pusat Statistik Kabupaten dalam angka di setiap kabupaten Provinsi Sulawesi Barat tahun 2023.

BAB II

KAJIAN PUSTAKA

2.1 Analisis Multivariat

Menurut Mulyani *et al.*, (2022), analisis multivariat didefinisikan sebagai metode pengolahan variabel dalam jumlah yang banyak. Selain itu, analisis multivariat juga merupakan suatu metode analisis yang digunakan pada metode statistika dengan tujuan penggunaannya adalah untuk menganalisis data yang terdiri dari variabel dalam jumlah yang banyak serta diduga antar variabel tersebut memiliki hubungan satu sama lain.

Teknik analisis multivariat secara umum dapat dibagi menjadi dua kelompok besar yaitu :

1. *Dependence Methods* adalah teknik multivariat yang menggunakan satu variabel atau set variabel terikat sebagai dependen dan satu variabel lain sebagai variabel bebas atau independen.
2. *Interdependence Methods* adalah teknik multivariat yang melihat semua variabel secara bersamaan, tidak ada variabel yang didefinisikan sebagai bebas atau terikat.

2.2 Analisis Cluster

Analisis *cluster* didefinisikan sebagai metode analisis statistika yang didasarkan pada bagaimana masing-masing objek memiliki karakteristik yang mirip dan dapat digunakan untuk menempatkan kumpulan objek ke dalam dua grup bahkan lebih. Analisis *cluster* adalah metode analisis yang bertujuan untuk memilih objek dalam berbagai kelompok, dengan perbedaan karakteristik antara kelompok satu dan kelompok lainnya berbeda. Dalam analisis *cluster*, tiap kelompok bersifat *homogen* antar anggota dalam kelompok atau variasi objek dalam satu kelompok yang terbentuk sekecil mungkin. Solusi *cluster* sebagian besar tergantung pada variabel yang digunakan untuk mengevaluasi kesamaan (Ulinnuha & Veriani, 2020). Hasil analisis *cluster* dapat diubah dengan menambah atau mengurangi variabel yang terkait. Sedangkan menurut Faisal & Rifai (2023), pengelompokan

objek-objek yang memiliki karakteristik yang sama ke dalam satu kelompok dan antar kelompok berbeda satu sama lain dapat disebut sebagai *cluster*.

Cluster memiliki tingkat *homogenitas* (kesamaan) yang tinggi antara anggota di dalamnya dan tingkat *heterogenitas* (perbedaan) yang tinggi antara anggota di luarnya. Beberapa istilah penting dalam analisis *cluster* menurut (Talakua *et al.*, 2017), termasuk:

1. *Distances between cluster centers*, yaitu jarak yang menunjukkan bagaimana terpisahnya pasangan individu *cluster*.
2. *Cluster membership*, yaitu keanggotaan yang menunjukkan *cluster* untuk setiap objek yang menjadi anggotanya.
3. *Cluster centers*, merupakan titik awal di mulai pengelompokan dalam *cluster non-hierarki*.
4. *Cluster centroid*, yaitu nilai rata-rata variabel dari semua objek atau observasi dalam *cluster* tertentu.
5. *Agglomeration schedule*, merupakan jadwal yang memberikan informasi tentang objek atau kasus yang dikelompokkan pada setiap tahap pada suatu proses analisis *cluster* yang *hierarki*.

2.3 Metode *Hierarchical Cluster*

Hierarchical cluster terbagi atas dua metode yaitu *Agglomerative* (penggabungan) dan *divisive* (pemecahan). Pada metode *Agglomerative* setiap penelitian pada awalnya diasumsikan sebagai *cluster* tersendiri dimana terdapat kelompok sesuai jumlah penelitiannya. Selanjutnya untuk dua *cluster* yang memiliki kesamaan terdekat digabungkan menjadi *cluster* baru, dengan begitu jumlah *cluster* akan berkurang satu untuk tiap tahapan. Sedangkan pemecahan, akan dimulai dengan satu *cluster* besar yang mengandung semua informasi, dan kemudian informasi yang berbeda akan dipisahkan dan dibentuk kelompok yang lebih kecil. Proses ini akan berlanjut, sampai setiap observasi menghasilkan kelompok terpisah. Setelah itu, perhitungan akan diteruskan ke objek lain dengan kedekatan berikutnya atau kedua. Kemudian dilanjutkan hingga kelompok

membentuk dendrogram atau pohon, dengan tingkatan yang jelas antara objeknya (Ulinnuha & Veriani, 2020).

Secara umum, algoritma *cluster Aggolmerative Hierarchical Clustering* untuk mengelompokkan N objek adalah sebagai berikut (Rachmatin, 2014).

1. Mulai dengan N cluster, setiap cluster mengandung unsur tunggal dan sebuah matriks simetris $D : \{d_{jl}\}$
2. Tentukan jarak terhadap pasangan cluster terdekat
3. Menggabungkan masing-masing cluster yang telah ditunjukkan memiliki jarak yang relatif dekat.
4. Ulangi langkah 2 dan 3 sebanyak $(n-1)$ kali, sampai semua objek akan berada dalam cluster tunggal.

2.3.1 Agglomerative Hierarchical Clustering

Agglomerative Hierarchical Clustering merupakan metode pengelompokan hierarki dengan pendekatan *bottom-up*, yaitu pasangan objek terhubung secara berurutan untuk menghasilkan cluster yang lebih besar (Kasoqi *et al.*, 2021). Dalam *Aggolmerative Hierarchical Clustering* terdiri dari enam metode diantaranya, metode *Single Linkage*, metode *Complete Linkage*, metode *Average Linkage*, metode *Centroid*, metode *Ward*, dan metode *Median Linkage*. Dalam penelitian ini hanya akan berfokus menggunakan metode *Average Linkage*, *Ward* dan *Median Linkage*.

2.3.1.1 Average Linkage

Average Linkage adalah metode yang mengelompokkan dua objek berdasarkan jarak rata-rata dengan meminimumkan rata-rata jarak antar pasangan (Tias & Primandari, 2023). *Average Linkage* juga diartikan sebagai pengelompokan dua objek dengan jarak terdekat pertama dengan terdekat kedua dan seterusnya adalah cara kerja pada metode ini. Pada dasarnya metode ini bekerja dengan meminimumkan rata-rata jarak antar cluster. Adapun untuk menghitung jarak dua cluster pada metode *Average Linkage* yang merupakan jarak rata-rata dapat dihitung menggunakan persamaan berikut (Widyadhan *et al.*, 2021).

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)} N_W} \quad (2.1)$$

dengan,

d_{ik} : jarak antar objek i pada *cluster* (UV) dan objek k pada *cluster* W .

$N_{(UV)}$: banyaknya anggota dalam *cluster* UV .

N_W : banyaknya anggota dalam *cluster* W .

2.3.1.2 Metode Ward

Metode *Ward* didefinisikan sebagai metode pengelompokan dengan menggunakan perhitungan yang komprehensif dan memaksimalkan kesamaan dalam satu kelompok (Tias & Primandari, 2023). Pengertian analisis *cluster* dengan metode *Ward* juga merupakan metode yang berusaha untuk meminimalkan variasi antar objek yang ada dalam satu *cluster* (Lesnussa *et al.*, 2021). Metode *Ward* dapat dihitung berdasarkan persamaan rumus dalam menghitung *Sum Of Square* (SSE) sebagai berikut:

$$SSE = \sum_{i=1}^n (x_i - \bar{x})'(x_i - \bar{x}) \quad (2.2)$$

dengan,

x_i : vektor objek ke- i

$\bar{x}$: vektor rata-rata semua objek dalam *cluster*

2.3.1.3 Metode Median Linkage

Median *Linkage* merupakan salah satu metode dalam pengelompokan yang berdasarkan nilai mediannya. Metode ini juga digunakan untuk menghindari bobot vektor rata-rata yang diubah berdasarkan ukuran *cluster* (Aisyi, 2024). Berikut adalah rumus yang digunakan dalam *Median Linkage* adalah sebagai berikut:

$$d_{mj} = \sqrt{\frac{d_{kj}^2}{2} + \frac{d_{lj}^2}{2} + \frac{d_{kl}^2}{4}} \quad (2.3)$$

dengan,

d_{mj} : jarak antara *cluster* m dan j

m : gabungan kelompok yang terdiri dari *cluster* k dan l , dengan $m = (k,l)$

d_{kj} : jarak antara *cluster* k dan j

d_{lj} : jarak antara *cluster* l dan j

d_{kl} : jarak antara *cluster* k dan l

2.3.2 Divisive

Metode ini berlawanan dengan metode *Agglomerative*. Dimulai dengan satu *cluster* besar yang mencakup keseluruhan objek. Selanjutnya, objek yang memiliki ketidakmiripan secara besar dipisahkan sehingga membentuk *cluster* yang lebih kecil. Pemisahan ini dilanjutkan sampai tercapai sejumlah *cluster* yang diinginkan. Salah satu metode *Divisive* yang termasuk dalamnya adalah metode jarak rata-rata *splinter*. Metode ini didasarkan pada perhitungan jarak rata-rata masing-masing objek dengan objek pada *cluster splinter* (terpisah) dan jarak rata-rata objek dengan objek lain pada *cluster*-nya. Proses dimulai dengan memisahkan dua objek dengan jarak terjauh sehingga terbentuk dua *cluster*. Kemudian, lakukan perbandingan jarak rata-rata masing-masing objek ke *cluster splinter* dan *cluster*-nya sendiri. Apabila suatu objek lebih dekat ke *cluster splinter* daripada ke *cluster*-nya sendiri, maka objek tersebut harus dikeluarkan dari *cluster*-nya dan dipindahkan ke *cluster splinter* apabila komposisinya sudah stabil, yaitu jarak suatu objek ke *cluster*-nya sudah sama dengan jarak suatu objek ke *cluster*.

Tahapan dari pengelompokan *divisive analysis* adalah sebagai berikut (Kasoqi *et al.*, 2021):

1. Melakukan standarisasi data
2. Menghitung jarak *Euclidean*
3. Menghitung rata-rata setiap objek dengan objek lainnya berdasarkan jarak *Euclidean*
4. Menentukan objek yang memiliki nilai rata-rata yang terbesar dan berubah menjadi *splinter group*.

5. Hitung selisih nilai antara objek *splinter group* dengan nilai rata-rata setiap objek yang tersisa.
6. Tentukan objek yang memiliki nilai selisih terbesar antara objek *splinter group* dengan nilai rata-rata. Jika nilai selisih positif, maka objek dengan nilai selisih terbesar bergabung dalam *splinter group*.
7. Mengulang langkah (4) sampai (7) sedemikian sehingga semua nilai selisih antara elemen matriks *splinter group* dengan nilai rata-rata bernilai *negative* dan *cluster* terbagi menjadi dua *cluster* baru.

2.4 Jarak Euclidean

Ukuran jarak yang digunakan dalam *cluster* pada metode *Avarage Linkage*, *Ward* dan *Median Linkage* adalah jarak *Euclidean*. Ukuran jarak adalah ukuran ketidak miripan yang biasa digunakan untuk data skala metrik, dimana kesamaan yang lebih sedikit muncul untuk jarak yang lebih besar, sedangkan kesamaan yang lebih kecil menunjukkan bahwa suatu objek akan lebih mirip dengan objek lainnya. Sehingga untuk penelitian ini menggunakan ukuran jarak *Euclidean*. Adapun rumus dari jarak *Euclidean* adalah sebagai berikut (Lesnussa *et al.*, 2021).

$$d_{i,u} = \sqrt{\sum_k^n (x_{ik} - x_{jk})^2} \quad (2.4)$$

dengan,

$d_{i,u}$: jarak antara objek ke- i dan objek ke- j

n : jumlah variabel *cluster*

x_{ik} : data dari objek ke- i pada variabel ke- k

x_{jk} : data dari objek ke- j pada variabel ke- k

k : data dari objek ke-1, 2, ..., p

Jarak *Euclidean* mempunyai kelebihan dibandingkan dengan metode perhitungan lain. Kelebihan dalam perhitungan jarak ini adalah sebagai berikut (Pribadi *et al.*, 2022):

1. Perhitungan jarak menggunakan *Euclidean* lebih umum digunakan.
2. *Euclidean* mempunyai hasil yang lebih optimal dibandingkan dengan perhitungan yang lainnya.

3. Referensi dari *Euclidean* lebih banyak karena secara umum perhitungan *Euclidean* lebih sering digunakan dalam perhitungan jarak pada *Clustering*.

2.5 Koefisien Korelasi *Chopenetic*

Koefisien korelasi *chopenetic* adalah koefisien korelasi antara elemen asli matriks ketidakmiripan (Matriks jarak kuadrat *Euclidean*) dan elemen yang dibuat oleh dendrogram. Sehingga korelasi ini adalah salah satu ukuran yang dapat digunakan untuk menguji validitas hasil *clustering* (Mahadesyawardani *et al.*, 2024). Menurut Mahadesyawardani *et al.*, (2024) rumus berikut dapat digunakan untuk menghitung koefisien korelasi *chopenetic*:

$$r_{coph} = \frac{\sum_{i < j} (d_{ij} - \bar{d})(d_{c_{ij}} - \bar{d}_c)}{\sqrt{\left[\sum_{i < j} (d_{ij} - \bar{d})^2 \right] \left[\sum_{i < j} (d_{c_{ij}} - \bar{d}_c)^2 \right]}} \quad (2.5)$$

dengan,

r_{coph} : koefisien korelasi *chopenetic*

d_{ij} : jarak *squared euclidean* antara objek i dan j

$\bar{d}$: rata-rata jarak *squared euclidean* antara objek i dan j

$d_{c_{ij}}$: jarak *chopenetic* antara objek i dan j

$\bar{d}_c$: rata-rata jarak *chopenetic* antara objek i dan j

1. Nilai koefisien korelasi *chopenetic* berkisar antara -1 dan 1, nilai r_{coph} mendekati 1 berarti solusi yang dihasilkan dari proses *clustering* baik.
2. Nilai 1 menunjukkan bahwa dendrogram adalah representasi jarak asli yang sempurna.
3. Nilai 0 menunjukkan bahwa tidak ada hubungan antara dendrogram dan jarak asli.
4. Nilai -1 menunjukkan bahwa dendrogram berlawanan arah dengan jarak.

Semakin tinggi nilai korelasi *chopenetic* mendekati nilai 1 maka semakin baik dendrogram dalam mempresentasikan struktur data asli. Sebaliknya jika semakin rendah nilai *chopenetic* maka semakin kurang tepat dendrogram dalam mempresentasikan struktur data asli.

2.6 Validasi Cluster

Setelah mencapai hasil *cluster*, langkah selanjutnya adalah mengevaluasi kualitas *cluster* dengan memeriksa kemampuan *cluster* untuk membedakan data yang ada sesuai dengan variabel atau karakteristik dari subjek yang digunakan untuk *cluster* (Pratiwi *et al.*, 2019). Validasi *cluster* yang digunakan untuk menentukan jumlah *cluster* optimal terdiri dari 3 macam yaitu validasi internal, validasi stabilitas, dan validasi biologis. Pada penelitian ini, validasi *cluster* internal dan validasi stabilitas digunakan untuk memvalidasi hasil *cluster* diantaranya: Indeks *Connectivity*, Indeks *Silhouette*, Indeks *Dunn*, dan metode WSS (Mahadesyawardani *et al.*, 2024).

2.6.1 Indeks *Connectivity*

Indeks *Connectivity* memiliki nilai diantara 0 sampai tak hingga. Semakin kecil nilai Indeks *Connectivity* semakin baik *cluster* yang terbentuk. Rumus perhitungan Indeks *Connectivity* adalah sebagai berikut :

$$Conn(C) = \sum_{I=1}^N \sum_{j=1}^L x_{inn_{i(j)}} \quad (2.6)$$

dengan,

$Conn(C)$: Indeks *Connectivity*

$nn_{i(j)}$: Pengamatan tetangga terdekat dari data ke- i ke data ke- j

N : Banyak pengamatan

L : Banyak *cluster*

Jika nilai Indeks *Connectivity* semakin kecil maka mengindikasikan semakin baik banyak *cluster* yang terbentuk.

2.6.2 Indeks *Silhouette*

Indeks *Silhouette* dapat digunakan untuk menghitung tingkat keheterogenan anggota internal dan antar *cluster*. Koefisien *Silhouette* berkisar antara -1 dan 1. Nilai Koefisien *Silhouette* yang lebih mendekati satu menunjukkan hasil validasi banyaknya *cluster* yang terbentuk yang lebih baik dan sebaliknya, nilai Koefisien *Silhouette* yang lebih mendekati -1 menunjukkan hasil validasi banyaknya *cluster*

yang terbentuk yang lebih buruk. Berikut ini adalah cara menghitung nilai Koefisien *Silhouette*:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.7)$$

dengan :

$a(i)$: jarak rata-rata antara i dan seluruh pengamatan lainnya pada *cluster* yang sama.

$b(i)$: jarak rata-rata antara i dengan pengamatan *cluster* terdekat.

2.6.3 Indeks *Dunn*

Indeks *Dunn* adalah rasio dari jarak terkecil antara objek dalam *cluster* yang sama dengan jarak terbesar antara objek dalam *cluster* yang sama. Nilai Indeks *Dunn* yang lebih tinggi menunjukkan bahwa jumlah *cluster* yang terbentuk lebih optimal. Ditunjukkan dengan D , Indeks *Dunn* dapat dihitung dengan menggunakan rumus yang ditemukan dalam persamaan berikut:

$$D = \frac{d_{\min}}{d_{\max}} \quad (2.8)$$

dengan,

$d_{\min}$: jarak terkecil antara objek pada *cluster* yang berbeda

$d_{\max}$: jarak terbesar antar objek pada *cluster* yang sama

2.6.4 Metode WSS (*Within-cluster Sum of Squares*)

Metode WSS adalah salah satu metode yang dapat digunakan dalam menentukan *cluster* paling optimal dari sebuah data. Metode ini dimulai dengan $k = 2$, dan terus meningkat dalam setiap langkah dengan ditambah 1 pada nilai k . Misalnya pada nilai $k = 3$, jika terjadi perubahan drastis yang berbanding terbalik dengan nilai sebelumnya, maka nilai sebelum terjadinya perubahan tersebut dianggap sebagai jumlah *cluster* yang paling optimal (Arif *et al.*, 2020). Adapun rumus WSS adalah sebagai berikut:

$$WSS = \sum_{l=1}^k \sum_{i \in S_k} \sum_{j=1}^p (x_{lj} - \bar{x}_{kj})^2 \quad (2.9)$$

dengan,

S_k : pengamatan di *cluster k*

$\bar{x}_{kj}$: rata- rata objek *j* dari *cluster* pusat untuk *cluster k*

p : jumlah total dimensi data

2.7 Produksi Pertanian

Pertanian adalah suatu jenis kegiatan produksi yang berlandaskan pada proses pertumbuhan dari tumbuh-tumbuhan dan hewan. Sedangkan, produksi merupakan suatu kegiatan untuk menciptakan/menghasilkan atau menambah nilai guna terhadap suatu barang atau jasa untuk memenuhi kebutuhan setiap orang atau badan (produsen). Sehingga produksi hasil pertanian dapat diartikan sebagai proses kombinasi dan koordinasi material-material dan kekuatan-kekuatan dalam pembuatan suatu barang atau jasa yang disebut *output* atau produk (Mukuan, *et al.*,2022).

Berdasarkan Badan Pusat Statistik Provinsi Sulawesi Barat, sektor pertanian merupakan salah satu sektor yang vital di kehidupan manusia. Sektor tersebut memiliki kontribusi yang sangat signifikan terhadap pencapaian tujuan program *Sustainable Development Goals* (SDG's) kedua, yaitu tidak ada kelaparan, mencapai ketahanan pangan, perbaikan nutrisi, serta mendorong budidaya pertanian yang berkelanjutan. Peran sektor pertanian di Indonesia juga menjadi sangat penting karena merupakan penyumbang terbesar ketiga terhadap Produk Domestik Bruto (PDB) yang berperan sebagai pendorong pertumbuhan ekonomi nasional.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan 2 poin berikut:

1. Metode *Average Linkage* merupakan metode *Agglomerative Hierarchical Clustering* terbaik dibandingkan dengan dua metode lainnya, yaitu metode *Ward* dan metode *Median Linkage*, dalam mengelompokkan kecamatan-kecamatan di Provinsi Sulawesi Barat berdasarkan hasil produksi pertanian tahun 2023. Pemilihan metode terbaik ini didasarkan pada nilai koefisien korelasi *Chopenetic* tertinggi. Nilai koefisien korelasi *Chopenetic* untuk metode *Average Linkage* sebesar 0,9658, sedangkan untuk metode *Ward* dan *Median Linkage* masing-masing sebesar 0,7793 dan 0,9412. Dengan nilai *Chopenetic* tertinggi tersebut, metode *Average Linkage* menunjukkan kualitas pengelompokan yang paling representatif terhadap data asli.
2. Jumlah *cluster* optimal yang terbentuk adalah dua *cluster* dengan karakteristik berbeda. *Cluster* 1 pada komoditas variabel durian, kopi, kakao, terung, jahe, bayam, dan rambutan memiliki karakteristik tinggi dibandingkan dengan komoditas tanaman yang serupa di *cluster* 2. Sedangkan pada *cluster* 2, hanya terdiri dari Kecamatan Banggae Timur, menunjukkan karakteristik hasil produksi pertanian yang menonjol pada variabel ubi kayu, kelapa, cabai besar, cabai rawit, lengkuas, mangga, pisang, dan pepaya, dibandingkan kecamatan-kecamatan lain di Sulawesi Barat.

5.2 Saran

Berdasarkan hasil dan pembahasan yang telah diperoleh, sehingga saran yang penulis berikan adalah sebagai berikut:

- a. Untuk peneliti selanjutnya, yang akan meneliti terkait hal yang serupa, kiranya dapat menambah variabel baru, menggunakan metode lainnya, dan mengembangkan analisis statistika yang digunakan terkait *clustering*.
- b. Peneliti selanjutnya yang akan menggunakan data yang sama, agar dapat lebih memperhatikan kecamatan di *cluster* 2 dengan memberikan metode baru dalam mengatasi *outlier*.

DAFTAR PUSTAKA

- Aisyi, G. R., & Mardhotillah, B. (2024). *Penggunaan Koefisien Silhouette Untuk Perbandingan Metode Ward dan Median Linkage Dalam Pengelompokan Kabupaten/Kota Berdasarkan Indikator Kesehatan Masyarakat Provinsi Jambi*. *Journal of Statistics and Its Application*. No.1, Vol.4, 1-9, : <https://repository.unja.ac.id/68332/>.
- Andaresta, D. P., Retnowati, D., Fatmawati, A., & Purnomo, S. D. (2024, August). Analisis Faktor yang Mempengaruhi Ketahanan Pangan di Indonesia. *In Prosiding Seminar Nasional Unars*. No.1, Vol.3, 433-443, : <https://unars.ac.id/ojs/index.php/prosidingSDGs/article/download/4983/381>.
- Amelia, R., Sri, A., & Rahayu, Y. (2023). *Analysis Of Indonesia's One Data Policy Implementation: The Role Of The Central Statistics Agency As A Statistical Data Coach Analysis Of Satu Data Indonesia Policy Implementation: The Role Of Bps-Statistics Indonesia As A Statistical Data Coach*, No.1, Vol.31, 708-719. <https://jurnal.darmaagung.ac.id/index.php/jurnaluda/article/view/3115>.
- Arif, A., Rusgiyono, Salsabilla, B., A., & Hoyyi, A. (2020). *Pengelompokan Provinsi-Provinsi Di Indonesia Menggunakan Metode Ward (Studi Kasus: Produksi Tanaman Pangan di Indonesia Tahun 2018)*. *Jurnal Gaussian*, No.1, Vol.9, 112–121. <http://ejournal3.undip.ac.id/index.php/gaussian>.
- Badan Pangan Nasional. (2023). *Indeks Ketahanan Pangan: Statistik Publikasi Portal Data Badan Pangan Nasional*, Jakarta Selatan, diakses pada tgl 12 agustus 2024, : *Statistik Publikasi, Portal Data Badan Pangan Nasional*.
- Badan Pusat Statistik Provinsi Sulawesi Barat. (2024). *Potensi Pertanian Provinsi Sulawesi Barat*, BPS Provinsi Sulawesi Barat, Mamuju, diakses pada tgl 01 mei 2025.
- Badan Pusat Statistik Provinsi Sulawesi Barat. (2024). *Potensi Pertanian Provinsi Sulawesi Barat*, BPS Provinsi Sulawesi Barat, Mamuju, diakses pada tgl 12 agustus 2024.
- Badan Pusat Statistik Kabupaten Majene. (2024). *Kabupaten Majene dalam angka: Publikasi - Badan Pusat Statistik Kabupaten Majene*, diakses pada tgl 02 agustus 2024.
- Badan Pusat Statistik Kabupaten Mamuju. (2024). *Kabupaten Mamuju dalam angka*, Vol.36, 0215-4455, : *Publikasi - Badan Pusat Statistik Kabupaten Mamuju*, diakses pada tgl 02 agustus 2024.
- Badan Pusat Statistik Kabupaten Mamuju Tengah. (2024). *Kabupaten Mamuju Tengah dalam angka*, Vol.10, 2655-0423, : *Publikasi - Badan Pusat Statistik Kabupaten Mamuju Tengah*, diakses pada tgl 02 agustus 2024.

- Badan Pusat Statistik Kabupaten Mamasa. (2024). *Kabupaten Mamasa dalam angka*, Vol.15, 2087-5363, : *Publikasi - Badan Pusat Statistik Kabupaten Mamasa*, diakses pada tgl 02 agustus 2024.
- Badan Pusat Statistik Kabupaten Pasangkayu. (2024). *Kabupaten Pasangkayu dalam angka*, Vol.22, : *Publikasi - Badan Pusat Statistik Kabupaten Pasangkayu*, diakses pada tgl 02 agustus 2024.
- Badan Pusat Statistik Kabupaten Polewali Mandar. (2024). *Kabupaten Polewali Mandar dalam angka*, 2337-4810, : *Publikasi - Badan Pusat Statistik Kabupaten Polewali Mandar*, diakses pada tgl 02 agustus 2024.
- Faisal, A.S. & Rifai, N. A. K. (2023). Penerapan Metode Hierarchical Clustering Multiscale Bootstrap untuk Pengelompokan Indikator Indeks Pembangunan Manusia Tahun 2021 di Jawa Barat. *Bandung Conference Series: Statistics*, No.1, Vol.3, <https://doi.org/10.29313/bcss.v3i1.6327>.
- Kaufman & Rousseeuw, (2009). *Finding Groups in Data An Introduction to Cluster Analysis*, John Wiley & Sons, Inc., Hoboken, New Jersey : Canada.
- Kasoqi, I., Nor Hayati, M., Goejantoro, R., Program Studi Statistika, D., Matematika, J., & Mulawarman Alamat, U. (2021). *Pengelompokan Desa Atau Kelurahan Di Kutai Kartanegara Menggunakan Algoritma Divisive Analysis*, Vol. 9, Issue 2: <https://jurnal.unimus.ac.id/index.php/statistik/article/view/9251>.
- Latif., N.Z., (2024). *Pengelompokan Puskesmas di Kabupaten Polewali Mandar Tahun 2023 Berdasarkan Indikator Gizi Buruk Balita Dengan Model Agglomerative Hierarchical Clustering*, Skripsi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Program Sarjana Statistika, Universitas Sulawesi Barat, Majene.
- Lesnussa, Y. A., Talakua, M. W., & Matdoan, M. Y. (2021). Analisis Klaster untuk Pengelompokan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Pendidikan dengan Menggunakan Metode Ward. *Jurnal Statistika Dan Aplikasinya*, No.1, Vol.5, 51-60, <https://scholar.google.com/citations?user=9-z5er4AAAAJ&hl=id&oi=sra>.
- Mahadesyawardani, A., Zhafirab, A. A., Ariyawan, J., Humaira, E. P., Mardianto, M. F. F., Amelia, D., & Ana, E. (2024). Pengelompokan Kabupaten dan Kota di Jawa Timur berdasarkan Percepatan Pemulihan Ekonomi Menggunakan Pendekatan Hirearki. *Ekspansional*, No.1, Vol.15, 39. <https://doi.org/10.30872/ekspansional.v15i1.1273>.
- Mukuan, C. V., Pongoh, F. D., Komalig, H. A. H., Kunci, K., Produksi, :, Analisis, P., Utama, K., & Gerombol, A. (2022). Pengelompokan Kecamatan Di Kabupaten Minahasa Berdasarkan Data Hasil Produksi Pertanian Tahun 2019 dengan Menggunakan Analisis Komponen Utama dan Analisis Gerombol. No.1, Vol.11, 12-17, <https://ejournal.unsrat.ac.id/index.php/decartesian>.

- Mulyani, S. R., Paramarta, V., Yuliaty, F., Fitriana, F., Zulfikar, T., Kosasih, K., ... & Nurmartiani, E. (2022). *Buku Multivariat Terapan*. Bandung.
- Nur A, M., Nur, I. M., Wilayah, P., Di, K., Kendal, K., Potensi, B., Pertanian, S., Hanandya, S., & Haris, A. (2021). Perbandingan Metode Single Linkage dan Average Linkage Pada Comparison Of Single Linkage and Average Linkage In The Grouping Of Sub-Districts In Kendal Based On The Potential Of Agriculture Sector, No.1, Vol.4, 2654-3257. <https://prosiding.unimus.ac.id/index.php/semnas/article/download/771/782>.
- Oktavianty, E., Lilies Handayani, dan, Studi Statistika, P., Matematika, J., Mipa, F., Tadulako, U., Soekarno Hatta Km, J., & Palu Sulawesi Tengah, T. (2019). *Pengelompokkan Kabupaten/Kota Di Sulawesi Berdasarkan Indikator Pendidikan Menggunakan Analisis Kluster Average Linkage Dan Median Linkage (District/City Clustering In Sulawesi Based On Education Indicators Using Average Linkage Cluster And Median Linkage Analysis)*, No.3, Vol.8, 191-197: <https://bestjournal.untad.ac.id/index.php/ejurnal/mipa/article/view/14960>.
- Pratiwi, S. I., Widiyarih, T., & Hakim, A. R. (2019). Analisis Kluster Metode Ward dan Average Linkage dengan Validasi Dunn Index dan Koefisien Korelasi Chopenetic (Studi Kasus: Kecelakaan Lalu Lintas Berdasarkan Jenis Kendaraan Tiap Kabupaten/Kota di Jawa Tengah Tahun 2018). *Jurnal Gaussian*, No.4, Vol.8, 486-495,: <https://ejournal3.undip.ac.id/index.php/gaussian/article/view/26747>.
- Pribadi, W. W., Yunus, A., & Sartika Wiguna, A. (2022). Perbandingan Metode K-Means Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang. In *Jurnal Mahasiswa Teknik Informatika* Vol. 6, Issue 2: <https://www.ejournal.itn.ac.id/index.php/jati/article/view/4808>.
- Rachmatin, D. (2014). Aplikasi Metode-Metode Agglomerative Dalam Analisis Kluster Pada Data Tingkat Polusi Udara. In *InfinityJurnal Ilmiah Program Studi Matematika STKIP Siliwangi Bandung*, Vol. 3, Issue 2. <https://ejournal.stkipsiliwangi.ac.id/index.php/infinity/article/view/59>.
- Talakua, M. W., Leleury, Z. A., & Taluta, A. W. (2017). Analisis cluster dengan menggunakan metode k-means untuk mengelompokkan Kabupaten/Kota di provinsi maluku berdasarkan indikator indeks pembangunan manusia Tahun 2014. *Barekeng: Jurnal Ilmu Matematika dan Terapan*, No.2, Vol.11, 119-128, : <https://ojs3.unpatti.ac.id/index.php/barekeng/article/view/353>.
- Tendean, T., & Purba, W. (2020). Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means. *Jurnal Sains dan Teknologi*, No.2, Vol.1, 5-11, : <https://ejournal.sisfokomtek.org/index.php/saintek/article/view/31>.

- Tias, F., & Primandari, A. H. (2023). Pengelompokan Kecamatan Di Kabupaten Merangin Berdasarkan Produksi Tanaman Perkebunan Tahun 2021 Menggunakan *Agglomerative Hierarchical Clustering*. In *Emerging Statistics and Data Science Journal* (Vol. 1, Issue 1). Emerging Statistics and Data Science Journal, No.1, Vol.1, 137-147, : <https://journal.uui.ac.id/esds/article/view/26982>.
- Ulinnuha, N., & Veriani, R. (2020). Analisis *Cluster* dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode *Complete Linkage, Average Linkage dan Ward*. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, No.1, Vol.5, 102-108, : <https://core.ac.uk/download/pdf/328277185.pdf>.
- Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 584–594. <https://journal.unnes.ac.id/sju/index.php/prisma/>.