

# METODE AVERAGE LINKAGE DAN WARD DALAM PENGELOMPOKAN KESEJAHTERAAN SUMATERA BARAT TAHUN 2021

Riska Ramadani<sup>1</sup>, Admi Salma<sup>2</sup>

<sup>1</sup> Prodi Statistika, Fakultas Matematika Ilmu Pengetahuan dan Alam Universitas Negeri Padang (UNP)

<sup>2</sup> Dosen Statistika, Fakultas Matematika Ilmu Pengetahuan dan Alam Universitas Negeri Padang (UNP)

## Article Info

### Article history:

Received January 30, 2022

Revised August 01, 2022

Accepted September 15, 2022

### Keywords:

Welfare

Average Linkage

Ward

Validity Cluster

### Kata Kunci:

Kesejahteraan

Average Linkage

Ward

Validitas Cluster

## ABSTRACT

Welfare in West Sumatra Province is still an important problem that is being faced the whole community in 2021. The government is still trying to improve performance the welfare of the community from all aspects. It is necessary to grouping the level of community welfare in the Province of West Sumatra by Regency/City using the average linkage and Ward methods. The results of the cluster analysis obtained by both methods show that the clusters formed are same but objects in the clusters are different. Consists of 3 clusters, namely cluster 1 with a low level of welfare, cluster 2 with a medium level of welfare and cluster 3 with a high level of welfare. Of the two methods, the average linkage method is better in grouping than Ward method the levels of welfare based on validity because it has a large Dunn Index value is 0,88 and a small Connectivity value is 5,85

## ABSTRAK

Kesejahteraan di Provinsi Sumatera Barat masih menjadi masalah penting sedang dihadapi oleh seluruh masyarakat di 2021. Dampak pandemi Covid-19 yang terjadi pada awal 2020 masih menyisakan persoalan yang belum tuntas terselesaikan hingga tahun 2021. Pemerintah masih berupaya meningkatkan kinerja untuk meningkatkan kesejahteraan masyarakat dari segala aspek. Oleh karena itu, perlu dilakukan pengelompokan tingkat kesejahteraan masyarakat di Provinsi Sumatera Barat menurut Kabupaten/Kota dengan menggunakan metode Average Linkage dan Ward. Hasil analisis cluster yang diperoleh kedua metode menunjukkan bahwa banyak cluster yang terbentuk sama namun anggota cluster pada kedua metode berbeda. Terdiri dari 3 cluster yaitu cluster 1 dengan tingkat kesejahteraan rendah, cluster 2 dengan tingkat kesejahteraan sedang dan cluster 3 dengan tingkat kesejahteraan tinggi. Dari kedua metode tersebut, metode Average Linkage lebih baik dalam mengelompokkan tingkat kesejahteraan dibandingkan dengan metode Ward berdasarkan validitas karena memiliki nilai Dunn Index yang besar yaitu 0,88 dan nilai Connectivity yang kecil yaitu 5,85.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## Riska Ramadani

Riska Ramadani

Prodi Statistika, Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,

Universitas Negeri Padang, Jl. Prof. Dr. Hamka, Air Tawar barat, Padang Utara, Padang, 25171

Padang, Sumatera Barat

Email: [ramadaniriska40@gmail.com](mailto:ramadaniriska40@gmail.com)

## 1. PENDAHULUAN

Kesejahteraan di Provinsi Sumatera Barat menjadi masalah penting yang dihadapi oleh pemerintah dan juga seluruh masyarakat. Dampak pandemi Covid-19 yang sudah terjadi di awal tahun 2020 masih menyisakan persoalan pelik yang belum juga terselesaikan dengan tuntas sampai tahun 2021. Pemerintah masih berusaha dalam meningkatkan kinerja untuk meningkatkan kesejahteraan sosial masyarakat dari segala aspek kehidupan yang meliputi daerah Kabupaten/Kota Provinsi Sumatera Barat. Berdasarkan data yang diperoleh dari Publikasi Badan Pusat Statistik (BPS) mengenai Kesejahteraan Sumatera Barat tahun 2021, kesejahteraan di Provinsi Sumatera Barat pada tahun 2021 dapat dilihat dari beberapa faktor utama yaitu sebagai berikut [1],[2].

Tabel 1. Kesejahteraan Provinsi Sumatera Barat Tahun 2020 dan 2021 Menurut Faktor Utama yang Mempengaruhi

Variabel	2020	2021
Jumlah Penduduk	5,534 Juta Jiwa	5,6 Juta Jiwa
Angka Harapan Hidup (AHH)	69,47 Tahun	69,59 Tahun
Persentase Keparahan Kemiskinan	0,92%	1,04%
Jumlah Penduduk Miskin	344,23 Ribu Jiwa	370,67 Ribu Jiwa

Sumber: BPS Sumatera Barat 2021 dan 2021

Pada Tabel 1 diperoleh informasi bahwa jumlah penduduk Sumatera Barat mengalami peningkatan pada tahun 2021. Peningkatan jumlah penduduk ini berdampak negatif terhadap meningkatnya persentase keparahan kemiskinan dan meningkatnya jumlah penduduk miskin yang mempengaruhi kesejahteraan masyarakat. AHH yang mengalami peningkatan sebesar 0,12 poin, bisa diartikan baik menunjukkan perbaikan kesehatan masyarakat kearah positif. Selain itu AHH berarti buruk karena semakin banyak bayi yang lahir, jumlah penduduk usia produktif semakin banyak maka masalah kesejahteraan sosial masyarakat baru bisa saja muncul dimasa mendatang. Permasalahan kesejahteraan masyarakat Sumatera Barat 2021 jika dilihat dari beberapa variabel utama yang mempengaruhi disebabkan oleh pandemi Covid-19 yang belum usai. Serta krisis perekonomian yang terjadi semenjak awal muncul Covid-19 berlanjut ke tahun 2021 dan juga dapat mempengaruhi kesejahteraan[3]. Dari persoalan-persoalan yang terjadi itulah jumlah penduduk miskin meningkat, taraf kehidupan serta harapan hidup masyarakat melemah. Kemiskinan yang terjadi akan menimbulkan berbagai persoalan lainnya ditengah masyarakat seperti kriminalitas yang meningkat [4].

Berdasarkan pemaparan masalah kesejahteraan yang terjadi di Provinsi Sumatera Barat tahun 2021 maka perlu dilakukan pengelompokan untuk mengetahui tingkat kesejahteraan masyarakat menurut objek yaitu Kabupaten/Kota. Salah satu analisis statistika yang digunakan untuk mengelompokkan objek adalah analisis *cluster*. Analisis *cluster* termasuk analisis multivariat teknik interdependen. Pada teknik interdependen tidak terdapat jenis variabel. Kedudukan semua variabel sama yaitu semua variabel merupakan variabel penjelas. Teknik ini digunakan untuk melihat keterkaitan hubungan antar semua variabel yang ada [5]. Analisis cluster adalah pengelompokan objek dari suatu gugus data pengamatan kedalam suatu kelompok, dimana objek yang memiliki karakteristik sama digabungkan kedalam satu *cluster*, kemudian objek yang berbeda karakteristiknya akan dimasukkan kedalam *cluster* lainnya. Analisis *cluster* terbagi dua (2) yaitu hirarki dan non hirarki. Metode hirarki terbagi 2 (dua) yaitu *agglomerative* (penggabungan) dan *divisive* (pemecahan)[6]. Pada penelitian ini digunakan analisis *cluster* hirarki *agglomerative*. Karena metode ini lebih sering digunakan dalam pengelompokan, kompleks dalam proses perhitungan dan memiliki berbagai metode sehingga dapat dibandingkan untuk memilih metode terbaik. Selain itu, metode ini tidak perlu menentukan jumlah *cluster* yang dibentuk diawal, namun dapat ditentukan berdasarkan dendogram yang terbentuk diakhir proses pengelompokan.



Metode *Average Linkage* merupakan metode analisis *cluster* hirarki *agglomerative* dimana pada metode ini proses pengelompokan dimulai dengan menghitung rata-rata antara dua objek. [7]. Kelebihan metode ini adalah dapat menggabungkan objek kedalam *cluster* dengan ragam yang kecil, memperhatikan struktur *cluster* yang terbentuk serta lebih stabil dibandingkan metode *single linkage* dan *Complete Linkage* [8].

Metode *Ward* yaitu metode *cluster* yang digunakan untuk mengelompokkan objek kedalam *cluster* dengan cara meminimumkan varians atau ragam yang ada dalam *cluster* berdasarkan nilai *Sum of Squares Error (SSE)*. Dua objek yang mempunyai *SSE* yang terkecil digabungkan menjadi satu *cluster* [7]. Kelebihan metode ini dapat membentuk *cluster* yang memiliki varians dalam *cluster* yang kecil sehingga antar objek dalam *cluster* memiliki homogenitas atau kesamaan yang tinggi [9].

Untuk menentukan perbandingan dua metode *cluster* dapat menggunakan nilai *Dunn Index (DI)* dan nilai *connectivity*. Nilai *DI* dan *connectivity* pada masing-masing metode dihitung kemudian dibandingkan. Nilai *DI* diharapkan besar dan nilai *connectivity* diharapkan kecil untuk menunjukkan metode *cluster* yang terbaik [10].

Rumusan masalah dalam penelitian ini yaitu (1) Bagaimana hasil pengelompokan tingkat kesejahteraan masyarakat Provinsi Sumatera Barat menurut Kabupaten/Kota Tahun 2021 menggunakan metode *Average Linkage* dan *Ward*?; (2) Metode mana yang terbaik dari metode *Average Linkage* dan *Ward* untuk pengelompokan tingkat kesejahteraan masyarakat Provinsi Sumatera Barat menurut Kabupaten/Kota Tahun 2021? Adapun tujuan dari penelitian ini yaitu (1) Untuk mengetahui hasil pengelompokan tingkat kesejahteraan masyarakat Provinsi Sumatera Barat menurut Kabupaten/Kota Tahun 2021 menggunakan metode *Average Linkage* dan *Ward* dan (2) Untuk mengetahui metode terbaik dari metode *Average Linkage* dan *Ward* dalam pengelompokan tingkat kesejahteraan masyarakat Provinsi Sumatera Barat menurut Kabupaten/Kota Tahun 2021.

## 2. METODE

Jenis penelitian ini adalah penelitian terapan (*applied research*). Penelitian terapan adalah penelitian yang bertujuan untuk menerapkan, menguji dan mengevaluasi suatu teori yang diterapkan untuk memecahkan masalah yang sedang terjadi atau yang akan terjadi dimasa depan. Sehingga hasil penelitian bisa direkomendasikan untuk diterapkan [11].

Jenis data yang digunakan adalah data sekunder. Sumber data yang digunakan dalam penelitian berasal dari publikasi Badan Pusat Statistik (BPS) mengenai indikator kesejahteraan Sumatera Barat tahun 2021.

Variabel yang digunakan untuk mengelompokan tingkat kesejahteraan masyarakat Sumatera Barat tahun 2021 yaitu jumlah penduduk, Angka Harapan Hidup, jumlah penduduk miskin, Pendapatan Domestik Regional Bruto (PDRB), Persentase rumah tangga menurut sumber air minum bersih, Persentase Angka Melek Huruf, Jumlah puskesmas. Kesejahteraan merupakan suatu keadaan yang mengacu kepada kehidupan ekonomi dan sosial masyarakat [11]. Kesejahteraan dapat diukur dari kondisi kesehatan, keadaan ekonomi, pendidikan, ketenagakerjaan, kemiskinan, kondisi konsumsi dan perumahan masyarakat, indikator Indeks Pembangunan Manusia (IPM) serta kebahagiaan dan kualitas hidup rakyat [12], [13]. Salah satu persoalan terkait dengan masalah jumlah penduduk adalah ketimpangan distribusi penduduk antar daerah Kabupaten dan Kota dan antara penduduk perdesaan dan perkotaan. Beberapa daerah kota penduduknya semakin besar sementara itu beberapa daerah kabupaten masih dihuni oleh penduduk yang relatif sedikit [1]. Kemiskinan yang merupakan masalah yang melibatkan banyak bidang serta memerlukan kebijakan dan program intervensi yang bersifat multidimensi pula agar kesejahteraan meningkat sehingga dapat terbebas dari masalah kemiskinan [14]. Angka Harapan Hidup adalah rata-rata prakira banyak tahun yang bisa dirasakan seseorang selama hidup. Adanya puskesmas bisa menjadi jaminan kehidupan yang sehat dan meningkatkan kesejahteraan seluruh masyarakat [2], [15]. Angka Melek Huruf (AMH) adalah perhitungan lama sekolah dalam tahun yang diharapkan bisa dirasakan anak pada umur tertentu. AMH dijadikan sebagai salah satu faktor kesejahteraan masyarakat bidang pendidikan [2]. Serta pendidikan rendah dijadikan salah satu faktor penentu kemiskinan [16]. Pendapatan Domestik Regional Bruto (PDRB) dapat digunakan untuk menentukan kesejahteraan disuatu daerah. Jika

PDRB tinggi maka dapat dikatakan masyarakat dalam daerah tersebut dalam keadaan sejahtera [2]. Penentu kesejahteraan masyarakat dapat juga dilihat dari Air minum yang bersih merupakan faktor kesejahteraan masyarakat paling penting karena dapat mewujudkan masyarakat yang sehat dan sejahtera [2],[17].

Proses dalam analisis *cluster* hirarki dapat dilakukan dengan standardisasi data. Proses ini dilakukan pada data yang memiliki perbedaan nilai yang cukup besar antar variabel. Perbedaan nilai yang cukup besar dapat mengakibatkan proses pengelompokan tidak valid. Oleh sebab itu perlu dilakukan proses standardisasi data dengan transformasi kedalam bentuk *z-skor* dengan rumus sebagai berikut [18].

$$z_i = \frac{x_i - \bar{x}}{s} \quad (1)$$

Setelah proses standardisasi maka proses analisis *cluster* dapat dilakukan.

Analisis *Cluster* merupakan analisis yang digunakan untuk mencari pola dalam gugus data dengan mengelompokkan objek atau pengamatan kedalam *cluster*. Tujuan analisis *cluster* ini yaitu untuk menemukan kelompok yang optimal dari objek atau pengamatan, dimana objek dalam setiap *cluster* memiliki karakteristik yang sama diukur menggunakan ukuran jarak dan antar *cluster* memiliki karakteristik yang berbeda [6]. Jarak yang digunakan dalam analisis *cluster* adalah jarak *Euclidean* sebagai berikut [6].

$$d_{ik} = \sqrt{\sum_{j=1}^p (x_{ij} - x_{kj})^2} \quad (2)$$

Penggunaan jarak *Euclidean* digunakan untuk metode *Average Linkage*. Metode ini menggunakan jarak antara dua *cluster* dimana masing-masing objek dianggap sebagai *cluster* masing-masing. Kemudian dua objek akan digabungkan menjadi satu *cluster* berdasarkan nilai jarak yang terkecil dan digabungkan dengan objek atau *cluster* yang lain dengan perhitungan jarak rata-rata. Metode ini dimulai dengan mencari matriks jarak  $D = \{d_{ik}\}$  untuk memperoleh objek-objek yang paling dekat (yang memiliki kemiripan) misalnya U serta V. Objek-objek ini digabungkan sehingga akan terbentuk *cluster* UV. Selanjutnya dihitung jarak antara *cluster* UV dengan *cluster* W, dengan menggunakan rumus sebagai berikut [7].

$$d_{(UV)W} = \frac{\sum_{i=1}^n \sum_{k=1}^n d_{ik}}{n_{(UV)}n_W} \quad (3)$$

Keterangan:

$d_{(UV)W}$  : Jarak antara *cluster* UV dengan *cluster* W

$d_{ik}$  : Jarak antara objek *i* pada *cluster* (UV) dengan objek *k* pada *cluster* W

$n_{(UV)}$  : Jumlah objek pada *cluster* (UV)

$n_W$  : Jumlah objek pada *cluster* W

Selanjutnya pada proses *cluster Ward* dilakukan dengan menggunakan jarak *Squared Euclidean* yang merupakan kuadrat dari jarak *Euclidean* dengan rumus sebagai berikut.

$$d_{ik}^2 = \sum_{j=1}^p (x_{ij} - x_{kj})^2 \quad (4)$$

Keterangan:

$d_{ik}$  : Jarak *Euclid* objek data ke-*i* dan objek ke-*k*

$p$  : Banyak peubah yang digunakan

$x_{ij}$  : Objek data ke-*i* pada peubah ke-*k*

$x_{kj}$  : Objek data ke-*j* pada peubah ke-*k*

Metode *Average Linkage* digunakan ketika metode *Single Linkage* dan *Complete Linkage* memiliki keterbatasan saat menentukan ukuran jarak dalam menelompokkan karakteristik yang sama [19].

Metode *Ward cluster* yang dibentuk memiliki ragam dalam *cluster* sekecil mungkin karena jarak dua *cluster* didasarkan pada nilai *Sum of Square Error (SSE)*. Metode ini juga efektif sehingga sering digunakan untuk dibandingkan dengan metode hirarki lainnya [20]. Secara umum, *SSE* dinotasikan sebagai berikut:



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

Keterangan:

$x_j$  : Vektor kolom yang berisikan nilai objek  $j$  dengan  $j = 1, 2, \dots, N$ .

Hasil dari metode *Ward* dapat ditampilkan dalam bentuk dendrogram [3].

Selanjutnya diperoleh persamaan jarak antar objek yang merupakan rumus yang digunakan untuk menentukan jarak dengan metode *Ward* sebagai berikut [6].

$$\begin{aligned} I_{ik} &= SSE_{UV} = \frac{1}{2}(x_i - x_k)'(x_i - x_k) \\ &= \frac{1}{2}d^2(x_i, x_k) \end{aligned} \quad (5)$$

Keterangan:

$I_{ik}$  : Jarak antara objek  $i$  dan objek  $k$

$x_i, x_k$  : Nilai rata-rata objek  $i$  dan  $k$

$d^2(x_i, x_j)$  : Jarak *squared euclidean* antara objek  $i$  dan objek  $j$

$k$  :  $1, 2, \dots, p$

$p$  : Jumlah variabel *cluster*

$ik$  :  $1, 2, \dots, n$

Jarak antar objek  $UV$  dan  $W$  dengan metode *Ward* yaitu sebagai berikut:

$$I_{(UV)W} = \frac{n_U + n_W}{n_{UV} + n_W} I_{UW} + \frac{n_V + n_W}{n_{UV} + n_W} I_{VW} - \frac{n_W}{n_{UV} + n_W} I_{UV} \quad (6)$$

Keterangan:

$I_{(UV)W}$  : Jarak antara *cluster*  $UV$  dan  $W$

$I_{UW}$  : Jarak antara *cluster*  $U$  dan  $W$

$I_{VW}$  : Jarak antara *cluster*  $V$  dan  $W$

$I_{UV}$  : Jarak antara *cluster*  $U$  dan  $V$

$n_U, n_V, n_W$  : Banyaknya objek pada *cluster* ke  $U, V, W$

Metode ini juga stabil sehingga sering dibandingkan dengan metode hirarki lainnya [21]. Serta metode lebih baik dari metode hirarki *agglomerative* lainnya [22].

Setelah proses perhitungan *cluster* hirarki selesai. Hasil dari analisis *cluster* hierarki akan ditampilkan dalam bentuk dendogram. Dendogram merupakan diagram atau grafik pohon yang digunakan secara hirarki untuk mengelompokkan objek dalam analisis *cluster* [23]. Setelah itu, untuk mengetahui perbandingan dua metode yang memiliki cara kerja yang paling baik dapat ditentukan berdasarkan validasi *cluster*. Metode yang biasa digunakan yaitu *Dunn Index (DI)* dan *Connectivity*. Nilai *DI* diharapkan besar untuk menunjukkan bahwa metode *cluster* tersebut baik dimana *DI* menghitung nilai perbedaan antara satu *cluster* dengan *cluster* yang lain. *DI* dirumuskan sebagai berikut [24]. Nilai *DI* besar jika lebih dari 1 [25].

$$DI = \min_{i=1, \dots, k} \left\{ \min_{j=i+1, \dots, k} \left( \frac{d(c_i, c_j)}{\max_{m=1, \dots, k} \text{diam}(c_m)} \right) \right\} \quad (7)$$

Keterangan:

$DI$  : *Dunn Index*

$k$  : Jumlah kelompok

$d(c_i, c_j)$  : Jarak antara *cluster*  $i$  dan *cluster*  $j$

$\text{diam}(c_m)$  : Diameter *cluster*  $i$

Sedangkan nilai *Connectivity* adalah nilai yang digunakan untuk melihat homogenitas dalam *cluster*. Nilai *connectivity* berada antara nol sampai tak hingga. Nilai ini diharapkan sekecil mungkin untuk menunjukkan bahwa metode *cluster* tersebut baik dimana objek dalam satu *cluster* memiliki kesamaan yang tinggi. *Connectivity* dirumuskan sebagai berikut [24].

$$Conn = \sum_{i=1}^N \sum_{j=1}^L x_{ij} \quad (8)$$

$x_{ij}$  : Objek  $j$  dan objek  $i$  yang terdekat

$Conn$  : *Connectivity*

Interpretasi suatu *cluster* merupakan nilai yang digunakan untuk mengetahui karakteristik masing-masing *cluster* agar dapat menjelaskan bagaimana perbedaan yang terjadi dari setiap *cluster* secara relevan. Ukuran yang bisa digunakan untuk proses interpretasi ini adalah menghitung *centroid* atau rata-rata variabel yang merupakan karakteristik masing-masing objek pada setiap *cluster*. Rumus yang digunakan adalah sebagai berikut [5].

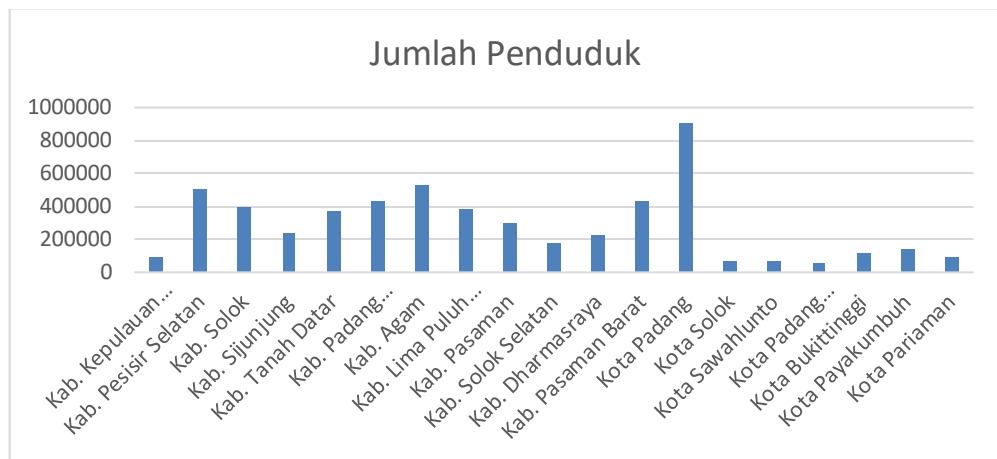
$$v = \frac{\sum_{i=1}^n y_i}{n} \quad (9)$$

Keterangan:

- $v$  : Nilai *centroid* atau nilai rata-rata  
 $y_i$  : Objek ke- $i$   
 $n$  : Banyaknya objek

### 3. HASIL DAN PAMBAHASAN

Berdasarkan hasil statistika deskriptif mengenai ringkasan data ditampilkan pada Gambar 1 sampai Gambar 7.



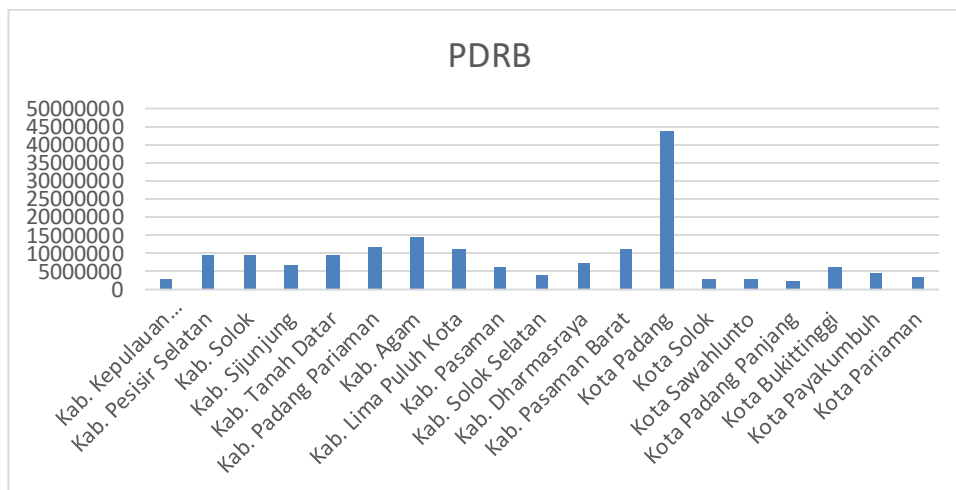
Gambar 1. Diagram Batang Jumlah Penduduk

Pada Gambar 1 diketahui jumlah penduduk terbanyak di Kota Padang sedangkan jumlah penduduk paling sedikit terdapat di Kota Padang Panjang. Banyaknya jumlah penduduk di Kota Padang membuat Kota ini memiliki jumlah penduduk miskin terbanyak. Sedangkan jumlah penduduk miskin paling sedikit terdapat pada Kota Sawahlunto. Hal tersebut menunjukkan bahwa kesejahteraan di Kota Padang masih menjadi masalah.



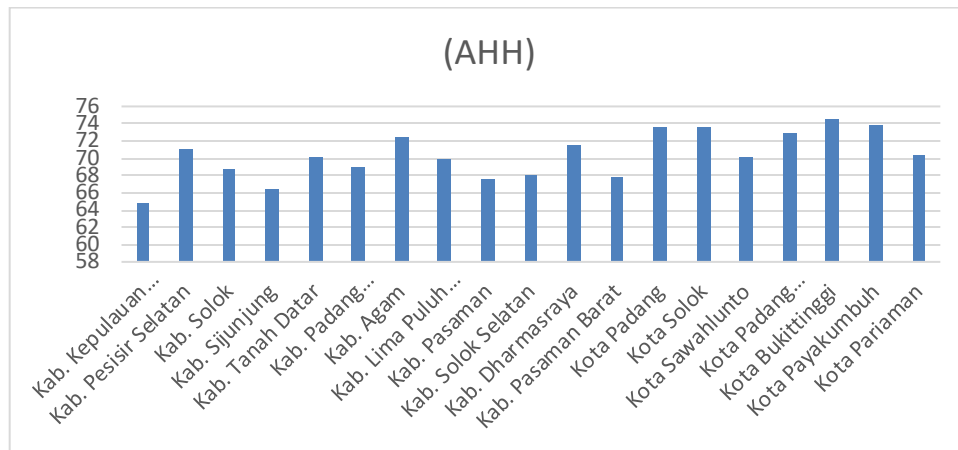
Gambar 2. Diagram Batang Jumlah Penduduk Miskin

Pada Gambar 2 diketahui bahwa Kota Padang memiliki jumlah penduduk miskin terbanyak dari Kabupaten/Kota lainnya yang ada di Provinsi Sumatera Barat. Banyaknya jumlah penduduk di Kota Padang membuat Kota ini memiliki jumlah penduduk miskin terbanyak. Sedangkan jumlah penduduk miskin paling sedikit terdapat pada Kota Sawahlunto. Hal tersebut menunjukkan bahwa kesejahteraan di Kota Padang masih menjadi masalah. Selanjutnya yaitu informasi mengenai Pendapatan Domestik Regional Bruto (PDRB) dan Angka Harapan Hidup (AHH) menurut Kabupaten/Kota.



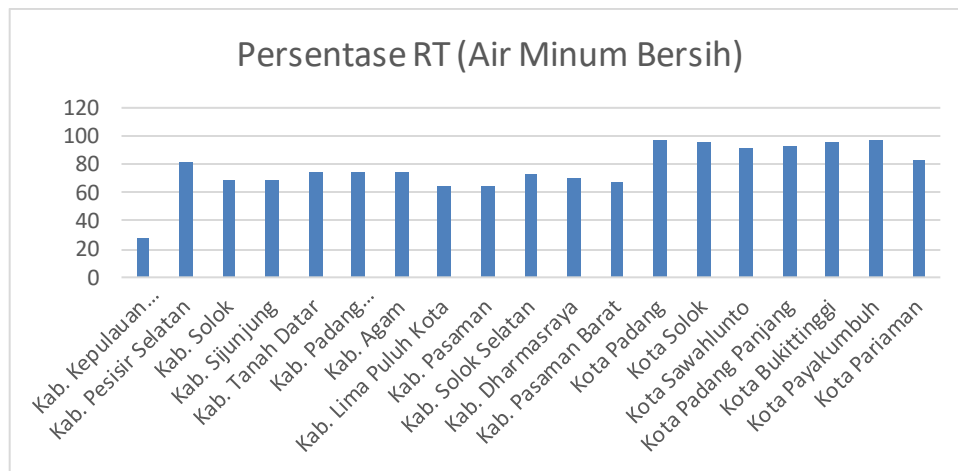
Gambar 3. Diagram Batang PDRB Menurut Kabupaten/Kota

Pada Gambar 3 diketahui bahwa PDRB paling tinggi berada pada Kota Padang. Kota Padang merupakan pusat pemerintahan di Provinsi Sumatera Barat. Kota ini menjadi lalu lintas kegiatan segala sektor seperti perekonomian, perdagangan, dan lainnya.



Gambar 4. Diagram Batang AHH Menurut Kabupaten/Kota

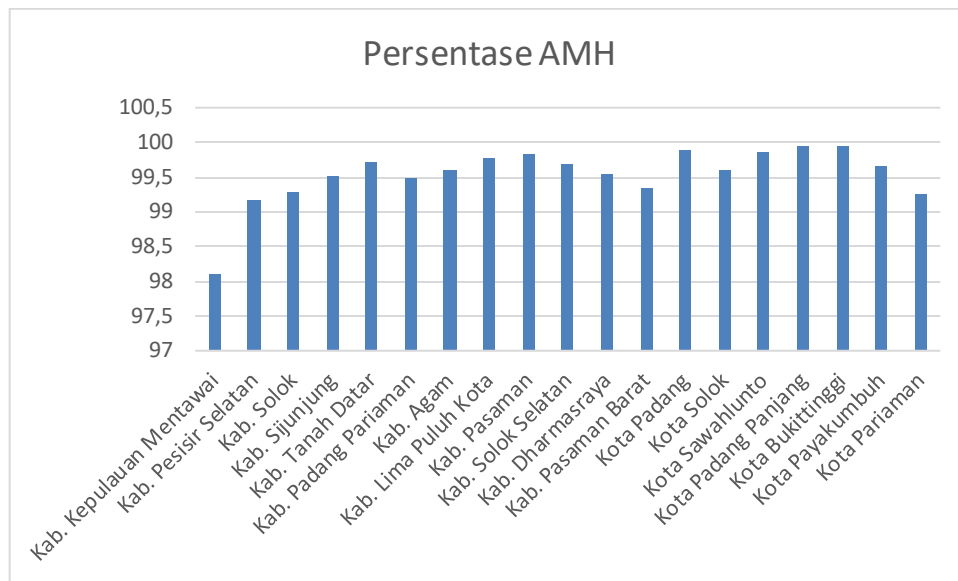
Sedangkan pada Gambar 4 diketahui bahwa Kota Bukittinggi memiliki rata-rata AHH lebih lama dibandingkan Kabupaten/Kota lainnya yang ada di Sumatera Barat. AHH yang tinggi ini menunjukkan bahwa kesejahteraan masyarakat dalam aspek kesehatan mengalami perbaikan ke arah yang lebih baik. Kabupaten Kepulauan Mentawai merupakan daerah dengan AHH terendah yang menunjukkan bahwa kesejahteraan masyarakat dalam aspek kesehatan masih jauh dari kata baik. Selanjutnya informasi yang terkandung pada variabel persentase air minum bersih, Angka Melek Huruf (AMH) dan jumlah puskesmas menurut Kabupaten/Kota.



Gambar 5. Diagram Batang Persentase Air Minum Bersih Menurut Kabupaten/Kota

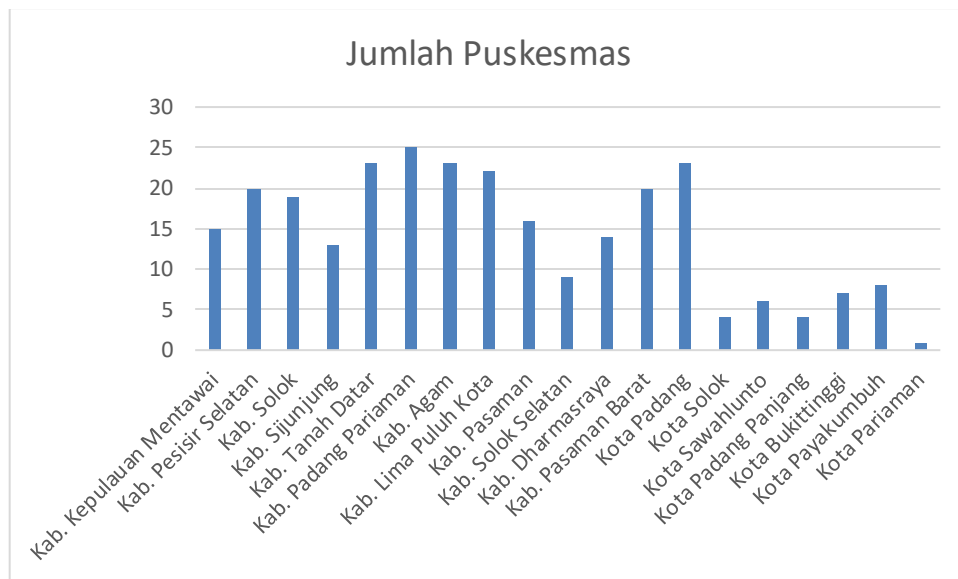
Gambar 5 diketahui bahwa Persentase rumah tangga dengan sumber air minum bersih paling tinggi terdapat di Kota Padang sedangkan persentase dengan sumber air minum bersih paling rendah adalah Kabupaten Kepulauan Mentawai. Hal ini menunjukkan kesejahteraan masyarakat dalam akses air minum bersih di Kota Padang lebih baik dibanding Kabupaten Kepulauan Mentawai. Karena Kota Padang memiliki akses yang lebih mudah dijangkau untuk penyaluran air bersih kepada masyarakat dibandingkan Kabupaten Kepulauan Mentawai.





Gambar 6. Diagram Batang AMH Menurut Kabupaten/Kota

Gambar 6 diketahui bahwa AMH yang terdapat di Kabupaten/Kota secara keseluruhan masyarakat sudah melek terhadap huruf dan adanya peningkatan terhadap kesejahteraan bidang pendidikan.



Gambar 7. Diagram Batang Jumlah Puskesmas Menurut Kabupaten/Kota

Pada Gambar 7 diketahui bahwa jumlah puskesmas paling banyak terdapat di Kabupaten Padang Pariaman. Jumlah puskesmas paling sedikit terdapat di Kota Solok. Jika semakin banyak jumlah puskesmas maka akses kesehatan bagi masyarakat lebih mudah.

Berikut ini proses pengelompokan objek tingkat kesejahteraan masyarakat Provinsi Sumatera Barat menurut Kabupaten/Kota tahun 2021 menggunakan metode *Average Linkage*. Metode pengelompokan ini dimulai dengan membentuk matriks jarak *Euclidean* dari data yang sudah distandardisasi untuk mencari kesamaan antar dua objek (Kabupaten/Kota) berdasarkan jarak terdekat. Berikut ini adalah matriks jarak *Euclidean*.

$$D = \begin{matrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 19 \end{matrix} & \begin{bmatrix} 0 & & & & & & \\ 5,43 & 0 & & & & & \\ 4,55 & 1,35 & 0 & & & & \\ 4,33 & 2,97 & 1,82 & 0 & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 4,93 & 3,62 & 3,11 & 2,28 & \dots & 0 \end{bmatrix} \end{matrix}$$

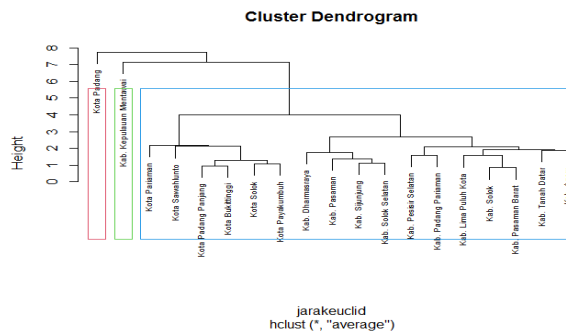
Setelah matriks jarak terbentuk, proses pengelompokan bisa dilakukan. Gabungkan dua objek menjadi satu *cluster* berdasarkan jarak *euclidean* yang paling kecil. Dari matriks jarak dapat diketahui bahwa jarak yang paling kecil adalah objek 3 (Kabupaten Solok) dan objek 12 (Kabupaten Pasaman Barat) dengan nilai jarak *euclidean* 0,56. Kedua objek tersebut bergabung menjadi satu *cluster*. Selanjutnya *cluster* yang telah terbentuk akan digabungkan dengan objek-objek lainnya dengan menggunakan jarak rata-rata metode *Average Linkage*. Misalnya akan dihitung jarak rata-rata antara objek 3,12 dengan objek 1 yaitu Kabupaten kepulauan Mentawai.

$$d_{(3,12)1} = \frac{d_{(1,3)} + d_{(1,12)}}{2}$$

$$d_{(3,12)1} = \frac{4,55 + 1,35}{2}$$

$$d_{(3,12)1} = 2,95$$

Proses pengelompokan berdasarkan jarak rata-rata ini akan berlanjut seterusnya semua objek berada pada *cluster* masing-masing. Hasil proses *cluster* metode *Average Linkage* dapat ditampilkan dalam bentuk dendrogram Gambar 8.



Gambar 8. Dendrogram *Average Linkage*

Dari Gambar 8 diketahui bahwa terbentuk tiga *cluster* sebagai berikut.

Tabel 2. Anggota *Cluster Average Linkage*

Cluster	Anggota
Cluster 1	Kabupaten Kepulauan Mentawai
Cluster 2	Kota Pariaman, Kota Sawahlunto, Kota Padang Panjang, Kota Bukittinggi, Kota Solok, Kota Payakumbuh, Kabupaten Dharmasraya, Kabupaten Pasaman, Kabupaten Sijunjung, Kabupaten Solok Selatan, kabupaten Pesisir Selatan, Kabupaten Padang Pariaman, Kabupaten Lima Puluh Kota, Kabupaten Solok, Kabupaten Pasaman Barat, Kabupaten Tanah Datar, Kabupaten Agam
Cluster 3	Kota Padang

Pada Tabel 2 dapat diketahui masing-masing dari anggota *cluster*. Pada *cluster* 1 dan *cluster* 3 anggota hanya terdiri dari 1 objek. Sedangkan *cluster* 2 terdiri dari 17 objek Kabupaten/Kota. Berikut ini hasil validasi metode *Average Linkage*.

Tabel 3. Validasi *Cluster Average Linkage*

<i>Dunn Index</i>	<i>Connectivity</i>
0,88	5,85

Dari Tabel 3 diketahui bahwa nilai *Dunn Index* dan nilai *Connectivity* dari metode *Average Linkage* dengan jumlah *cluster* tiga. Tabel 3 juga menunjukkan bahwa kesamaan objek dalam satu *cluster* sebesar 5,85. Sedangkan perbedaan antar satu *cluster* dengan *cluster* yang lain sebesar 0,88.

Interpretasi *cluster* yang terbentuk dengan menghitung rata-rata variabel pada setiap objek yang terdapat di setiap *cluster*. Hasil ini dapat memberikan gambaran mengenai karakteristik masing-masing *cluster* yang terbentuk dari proses pengelompokan dengan metode *Average Linkage*. Berikut ini adalah hasil rata-rata untuk interpretasi *cluster Average Linkage*.

Tabel 4 . Interpretasi *Cluster Metode Average Linkage*

Variabel	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
Jumlah Penduduk	87.623	266.930	909.040
AHH	64,7	70,5	73,7
Jumlah Penduduk Miskin	14.310	18.114	48.440
PDRB	2.956.008	7.284.773	43.631.433
Persentase Air Minum Bersih	27,2	78,9	97,9
Persentase AMH	98,1	99,6	99,9
Jumlah Puskesmas	15	14,1	23

Pada Tabel 4 menunjukkan karakteristik masing-masing *cluster*. Sehingga interpretasi *cluster* sebagai berikut.

1. *Cluster 1* dengan rata-rata variabel pada satu objek dalam *cluster* terendah. Sehingga *cluster 1* termasuk tingkat kesejahteraan rendah.
2. *Cluster 2* dengan rata-rata variabel pada tiap objek yang lebih tinggi dibandingkan *cluster 1*. Hal ini menunjukkan bahwa *cluster 2* adalah kelompok kesejahteraan sedang.
3. *Cluster 3* dengan rata-rata variabel pada satu objek tertinggi. Sehingga *cluster 3* termasuk tingkat kesejahteraan tinggi.

Selanjutnya adalah analisis data yang akan dilakukan pada pengelompokan tingkat kesejahteraan masyarakat Provinsi Sumatera Barat menurut Kabupaten/Kota tahun 2021 menggunakan metode *Ward*.

Metode pengelompokan ini dimulai dengan membentuk matriks jarak *Squared Euclidean* dari data yang sudah di standardisasi untuk mencari nilai *SSE*. Berikut ini adalah matriks jarak *Squared Euclidean*.

$$D^2 = \begin{matrix} 1 & \left[ \begin{array}{cccccc} 0 & & & & & \\ 29,48 & 0 & & & & \\ 20,70 & 1,82 & 0 & & & \\ 18,74 & 8,82 & 3,31 & 0 & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 19 & 24,30 & 13,10 & 9,67 & 5,19 & \dots & 0 \end{array} \right] \end{matrix}$$

Pengelompokan dengan metode *Ward* berdasarkan nilai *SSE* yang minimum. Berikut ini adalah nilai *SSE* antara 2 objek yang paling minimum yaitu objek 3 (Kabupaten Solok) dan objek 12 (Kabupaten Pasaman Barat).

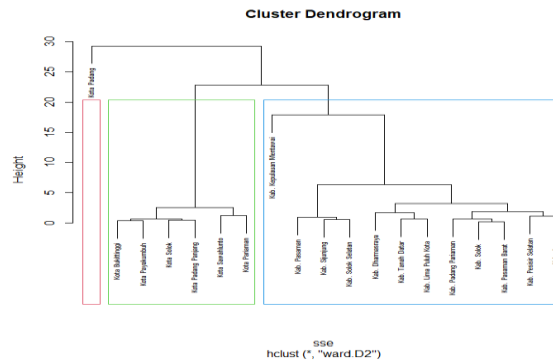
$$I_{(3,12)} = SSE_{(3,12)} = \frac{1}{2}(0,32) = 0,18$$

Kedua objek ini akan bergabung menjadi satu *cluster*. Selanjutnya pengelompokan dilakukan dengan *cluster* yang sudah terbentuk dengan objek-objek lainnya berdasarkan kenaikan nilai *SSE* yang paling minimum. Misalnya akan dihitung *cluster* yang telah terbentuk dengan objek 1 (Kabupaten Kepulauan Mentawai) sebagai berikut.

$$I_{(3,12)1} = \frac{1+1}{2+1}(29,48) + \frac{1+1}{2+1}(21,77) - \frac{1}{2+1}(0,18)$$

$$I_{(3,12)1} = 34,106$$

Proses pengelompokan berdasarkan nilai *SSE* ini akan berlanjut seterusnya semua objek berada pada *cluster* masing-masing. Hasil proses *cluster* metode *Ward* dapat ditampilkan dalam bentuk dendrogram Gambar 9.



Gambar 9. Dendrogram *Ward*

Dari Gambar 9 diketahui bahwa terbentuk tiga *cluster* sebagai berikut.

Tabel 5. Anggota Cluster *Average Linkage*

Cluster	Anggota
Cluster 1	Kabupaten Kepulauan Mentawai, Kabupaten Pasaman, Kabupaten Sijunjung, Kabupaten Solok Selatan, kabupaten Pesisir Selatan, Kabupaten Padang Pariaman, Kabupaten Lima Puluh Kota, Kabupaten Solok, Kabupaten Pasaman Barat, Kabupaten Tanah Datar, Kabupaten Agam, Kabupaten Dharmasraya
Cluster 2	Kota Padang
Cluster 3	Kota Pariaman, Kota Sawahlunto, Kota Padang Panjang, Kota Bukittinggi, Kota Solok, Kota Payakumbuh

Pada Tabel 5 dapat diketahui masing-masing dari anggota *cluster*. Pada *cluster* 1 terdiri dari 12 objek dan *cluster* 2 anggota hanya terdiri dari 1 objek. Sedangkan *cluster* 3 terdiri dari 6 objek Kabupaten/Kota. Berikut ini hasil validasi metode *Ward*.

Tabel 6. Validasi Cluster *Ward*

Dunn Index	Connectivity
0,27	8,57

Tabel 6 diketahui bahwa nilai *Dunn Index* dan nilai *Connectivity* dari metode *Ward* dengan jumlah *cluster* tiga. Tabel 6 juga menunjukkan bahwa kesamaan objek dalam satu *cluster* sebesar 8,57. Sedangkan perbedaan antar satu *cluster* dengan *cluster* yang lain sebesar 0,27.

Interpretasi *cluster* yang terbentuk dengan menghitung rata-rata variabel pada setiap objek yang terdapat di setiap *cluster*. Hasil ini dapat memberikan gambaran mengenai karakteristik masing-masing *cluster* yang terbentuk dari proses pengelompokan dengan metode *Ward*. Berikut ini adalah hasil rata-rata untuk interpretasi *cluster Ward*.

Tabel 7. Interpretasi Cluster Metode *Ward*

Variabel	Cluster 1	Cluster 2	Cluster 3
Jumlah Penduduk	339.643	909.040	91.619



AHH	69	73,7	72,6
Jumlah Penduduk Miskin	24.626	48.440	4.457
PDRB	8.721.140	43.631.433	3.690.578
Persentase Air Minum Bersih	67,6	97,9	92,9
Persentase AMH	99,4	99,9	99,7
Jumlah Puskesmas	18,2	23	6

Pada Tabel 7 menunjukkan karakteristik masing-masing *cluster*. Sehingga interpretasi *cluster* sebagai berikut.

1. *Cluster 1* dengan rata-rata variabel tiap objek dalam *cluster* terendah. Sehingga *cluster 1* termasuk tingkat kesejahteraan rendah.
2. *Cluster 2* dengan rata-rata variabel objek dalam *cluster* yang tertinggi. Hal ini menunjukkan bahwa *cluster 2* adalah kelompok kesejahteraan tinggi.
3. *Cluster 3* dengan rata-rata variabel tiap objek dalam *cluster* yang tinggi. Sehingga *cluster 3* termasuk tingkat kesejahteraan sedang.

Dari kedua metode pengelompokan diperoleh informasi bahwa metode pengelompokan kesejahteraan masyarakat di Provinsi Sumatera Barat lebih baik menggunakan metode *Average Linkage* lebih baik dibandingkan *Ward*. Hal ini dilihat dari nilai *DI* yang dihasilkan lebih besar yang menunjukkan perbedaan antar *cluster* yang terbentuk lebih tinggi dan nilai *Connectivity* yang dihasilkan kecil yang menunjukkan bahwa kesamaan objek dalam *cluster* yang terbentuk tinggi.

#### 4. Kesimpulan

Berdasarkan hasil yang diperoleh, terdapat 3 *cluster* yang terbentuk dalam pengelompokan tingkat kesejahteraan masyarakat di Provinsi Sumatera Barat menurut Kabupaten/Kota Tahun 2021 menggunakan metode *Average Linkage* dan *Ward*. Kedua metode menghasilkan tingkatan *cluster* yang sama namun anggota (objek) dalam *cluster* yang berbeda. Jumlah *cluster* adalah 3 dengan kategori tingkatan *cluster* yang terbentuk adalah rendah, sedang dan tinggi.

Dari pengelompokan tingkat kesejahteraan masyarakat di Provinsi Sumatera Barat menurut Kabupaten/Kota dengan menggunakan metode *Average Linkage* dan *Ward* dapat diketahui bahwa metode *Average Linkage* lebih baik dibandingkan metode *Ward*. Hal ini dapat diketahui dari validasi *cluster* menggunakan nilai *Dunn Index (DI)* dan nilai *connectivity*. Dari kedua metode pengelompokan yang digunakan dapat disimpulkan bahwa metode *Average Linkage* lebih baik dibandingkan metode *Ward*. Karena menghasilkan nilai *Dunn Index* yang lebih besar yaitu 0,88 yang berarti perbedaan antar satu *cluster* dengan *cluster* sangat tinggi. Serta nilai *Connectivity* yang lebih kecil yaitu 5,85 yang berarti kesamaan objek dalam satu *cluster* tinggi.

#### Ucapan Terimakasih

Ucapan syukur kepada Allah atas karunia-Nya penulis berhasil menyelesaikan jurnal ini. Tak lupa kepada orangtua yang selalu memberi do'a, dukungan moril serta materil kepada penulis. Sekaligus juga untuk dosen pembimbing yang telah memberikan arahan sehingga jurnal ini dapat terselesaikan.

#### REFERENSI

- [1] Badan Pusat Statistik, *Indikator Kesejahteraan Rakyat Provinsi Sumatera Barat 2021*. 2021.
- [2] Badan Pusat Statistik, "Indikator Kesejahteraan Rakyat Provinsi Sumatera Barat 2020," 2020.
- [3] S. Setyadi and L. Indriyani, "Dampak Pandemi Covid-19 Terhadap Peningkatan Anak," *J. Ekon. Dan Kebijak. Publik*, vol. 4, no. 1, pp. 53–66, 2021, [Online]. Available: <https://spektrumonline.com/2020/11/11/dampak-pandemi-covid-19-multidimensi/>.
- [4] H. Tarigan, J. H. Sinaga, and R. R. Rachmawati, "Dampak Pandemi Covid-19 Terhadap Kemiskinan di Indonesia," *Pus. Sos. Ekon. dan Kebijak. Pertan.*, pp. 457–479, 2020, [Online]. Available: <https://pse.litbang.pertanian.go.id/ind/pdf/files/23-BBRC-2020-IV-1-1-HLT.pdf>.
- [5] J. Hair, W. Black, B. Babin, and R. Anderson, *Multivariate Data Analysis*, 7th ed. Edinburgh: Pearson Education, 2014.
- [6] A. C. Rencher and W. F. Christensen, *Methods of Multivariate Analysis*. New Jersey: John Wiley & Sons, Inc, 2012.
- [7] R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis Richard Johnson Dean Wichern Sixth Edition*, 6th ed.

- United States of America: Pearson Education, 2014.
- [8] I. Ramadhani, Faiz dan Zain, "Analisis Regresi Logistik Biner Untuk Mengidentifikasi Faktor-Faktor Yang Mempengaruhi Status Penerimaan Beras Keluarga Miskin (Raskin) di Kecamatan Gunung Anyar," *Digit. Libr. ITS*, pp. 1–7, 2014.
- [9] G. Punj and W. Stewart, David, "Cluster Anaalysis in Marketing Research: Review and Suggestions for Application," *J. Mod. Afr. Stud.*, vol. 20, no. 2, pp. 134–148, 1983.
- [10] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan Tindakan*. Bandung: Alfabeta, 2013.
- [11] A. Sinaga, "Analisis Faktor-Faktor Yang Mempengaruhi Kesejahteraan Masyarakat Di Kota Medan (Studi Kasus Usaha Kecil Dan Menengah)," *J. Ilm. Methonomi*, vol. 2, no. 1, pp. 1–9, 2016.
- [12] R. A. Mulia and N. Saputra, "ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI," vol. 11, pp. 67–83, 2020.
- [13] I. G. dan I. N. K. Wiriana, "Analisis Faktor - Faktor Yang Mempengaruhi Kesejahteraan Masyarakat Di Kabupaten/Kota Provinsi Bali Tahun 2012 - 2018," *E-Jurnal EP Unud*, vol. 9[5], no. 3, pp. 1051–1081, 2018.
- [14] Edi Dores, "Istilah kemiskinan muncul ketika seseorang atau sekelompok orang tidak mampu mencukupi tingkat kemakmuran ekonomi yang dianggap sebagai kebutuhan minimal dari standar hidup tertentu. Kemiskinan dipahami sebagai keadaan kekura," *J. Econ. Econ. Educ.*, vol. 2, no. 2, p. 133, 2014, [Online]. Available: <https://media.neliti.com/media/publications/43009-ID-pengaruh-angka-melek-huruf-dan-angka-harapan-hidup-terhadap-jumlah-penduduk-misk.pdf>.
- [15] M. Ishaq, A. T. Rumiati, and E. O. Permatasari, "Analisis Faktor-Faktor yang Mempengaruhi Produksi Padi di Provinsi Jawa Timur Menggunakan Regresi Semiparametrik Spline," *J. Sains dan Seni ITS*, vol. 6, no. 1, pp. 1–5, 2017, doi: 10.12962/j23373520.v6i1.22451.
- [16] F. Anggadini, "Pendapatan Domestik Regional Bruto Perkapita Terhadap Kemiskinan Pada Kabupaten / Kota Di Provinsi Sulawesi Tengah Tahun 2010-2013," *e-Jurnal Katalogis*, vol. 3, no. 7, pp. 40–49, 2015.
- [17] Rohim and M. Triani, "Faktor – Faktor Yang Mempengaruhi Konsumsi Gas Di Indonesia," *J. Kaji. Ekon. dan Pembang.*, vol. 3, no. 2, p. 69, 2021, doi: 10.24036/jkep.v3i2.11594.
- [18] A. Kassambara, *Practical Guide To Cluster Analysis in R*, 1st ed. 2015.
- [19] O. Yim and K. T. Ramdeen, "Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data," *Quant. Methods Psychol.*, vol. 11, no. 1, pp. 8–21, 2015, doi: 10.20982/tqmp.11.1.p008.
- [20] B. Everitt, *Cluster analysis*, 5th ed. London: John Wiley & Sons, Inc., 2011.
- [21] P. Govender and V. Sivakumar, *Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)*, vol. 11, no. 1. Turkish National Committee for Air Pollution Research and Control, 2020.
- [22] S. Saracli, N. Dogan, and I. Dogan, "Comparison of hierarchical cluster analysis methods by cophenetic correlation," *J. Inequalities Appl.*, vol. 2013, no. December, 2013, doi: 10.1186/1029-242X-2013-203.
- [23] R. S. King, *Cluster Analysis and Data Mining: An Introduction*. Virginia: Mercury Learning and Information, 2015.
- [24] G. Brock, V. Pihur, S. Datta, and S. Dataa, "clValid: An R Package for Cluster Validation," *Solid State Commun.*, vol. 25, no. 4, pp. 1–22, 2008, doi: 10.1016/0038-1098(77)91248-0.
- [25] B. Rivera *et al.*, "Dunn's index for cluster tendency assessment of pharmacological data sets," *Can. J. Physiol. Pharmacol.*, vol. 90, no. 4, pp. 425–433, 2012, doi: 10.1139/Y2012-002.