

**PERBANDINGAN KINERJA METODE *AVERAGE LINKAGE*  
DAN *WARD* DALAM PENGELOMPOKAN TINGKAT  
KESEJAHTERAAN MASYARAKAT PROVINSI  
SUMATERA BARAT MENURUT KABUPATEN  
/KOTA TAHUN 2021**

**SKRIPSI**

*Diajukan Sebagai Salah Satu Persyaratan Guna Memperoleh Gelar  
Sarjana Statistika*



Oleh  
**RISKA RAMADANI**  
NIM 18337061

**PROGRAM STUDI SARJANA STATISTIKA  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2022**

**PERSETUJUAN UJIAN SKRIPSI**  
**PERBANDINGAN KINERJA METODE *AVERAGE LINKAGE* DAN *WARD***  
**DALAM PENGELOMPOKAN TINGKAT KESEJAHTERAAN**  
**MASYARAKAT PROVINSI SUMATERA BARAT MENURUT**  
**KABUPATEN/KOTA TAHUN 2021**

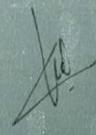
Nama : Riska Ramadani  
NIM : 18337061  
Program Studi : SI Statistika  
Jurusan : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 10 Februari 2022

Mengetahui,  
Ketua Jurusan Statistika

Disetujui Oleh  
Pembimbing

  
Dr. Dony Permana, M.Si.  
NIP. 19750127 200604 1 001

  
Admi Salma, S.Pd., M.Si.  
NIDN. 0025129003

**HALAMAN PENGESAHAN LULUS UJIAN SKRIPSI**

Nama : Riska Ramadani  
NIM : 18337061  
Program Studi : S1 Statistika  
Jurusan : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

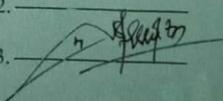
**PERBANDINGAN KINERJA METODE *AVERAGE LINKAGE* DAN *WARD*  
DALAM PENGELOMPOKAN TINGKAT KESEJAHTERAAN  
MASYARAKAT PROVINSI SUMATERA BARAT MENURUT  
KABUPATEN/KOTA TAHUN 2021**

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Tugas Akhir  
Program Studi S1 Statistika Jurusan Statistika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Negeri Padang

Padang, 10 Februari 2022

Tim Penguji  
Nania

1. Ketua : Admi Salma, S.Pd., M.Si
2. Anggota : Dodi Vionanda, M.Si., Ph.D
3. Anggota : Dr. Syafriandi, M.Si

1.   
2.   
3. 

## SURAT PERNYATAAN TIDAK PLAGIAT

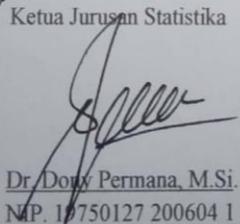
Saya yang bertanda tangan di bawah ini :

Nama : Riska Ramadani  
NIM/TM : 18337061/2018  
Program Studi : S1 Statistika  
Jurusan : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan, bahwa Skripsi saya dengan judul "**Perbandingan Kinerja Metode *Average Linkage* dan *Ward* dalam Pengelompokan Tingkat Kesejahteraan Masyarakat Provinsi Sumatera Barat Menurut Kabupaten/Kota Tahun 2021**" adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan. Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan Negara.

Demikianlah pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

✓ Diketahui oleh,  
Ketua Jurusan Statistika

  
Dr. Dony Permana, M.Si.  
NIP. 19750127 200604 1 001

Saya yang menyatakan,



Riska Ramadani  
NIM. 18337061

**PERBANDINGAN KINERJA METODE *AVERAGE LINKAGE*  
DAN *WARD* DALAM PENGELOMPOKAN TINGKAT  
KESEJAHTERAAN MASYARAKAT PROVINSI  
SUMATERA BARAT MENURUT KABUPATEN  
/KOTA TAHUN 2021**

**Riska Ramadani**

**ABSTRAK**

Kesejahteraan di Provinsi Sumatera Barat masih menjadi masalah penting yang sedang dihadapi oleh pemerintah dan juga seluruh masyarakat di Tahun 2021. Dampak pandemi Covid-19 yang terjadi pada awal tahun 2020 masih menyisakan persoalan pelik yang belum tuntas terselesaikan hingga tahun 2021. Pemerintah masih berupaya meningkatkan kinerja untuk meningkatkan kesejahteraan masyarakat dari segala aspek kehidupan. Oleh karena itu, perlu dilakukan pengelompokan tingkat kesejahteraan masyarakat di Provinsi Sumatera Barat menurut Kabupaten/Kota Tahun 2021 dengan menggunakan metode *Average Linkage* dan *Ward*. Metode *Average Linkage* mengelompokkan objek berdasarkan jarak terdekat menggunakan jarak *Euclidean* dan perhitungan jarak rata-rata. Sedangkan metode *Ward* mengelompokkan objek berdasarkan nilai *Sum of Squares Error (SSE)*.

Penelitian ini merupakan penelitian terapan yang dimulai dengan mempelajari teori-teori relevan dengan permasalahan. Data yang digunakan adalah data sekunder yang diperoleh dari *website* Badan Pusat Statistik (BPS) Sumatera Barat. Tujuan penelitian yaitu untuk mengelompokkan Kabupaten/Kota berdasarkan faktor-faktor yang mempengaruhi kesejahteraan masyarakat di Provinsi Sumatera Barat.

Hasil penelitian diperoleh bahwa dengan menggunakan metode *Average Linkage* dan *Ward* menunjukkan bahwa *cluster* yang terbentuk adalah sama. 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 dibandingkan dengan metode *Ward* dalam mengelompokkan tingkat kesejahteraan masyarakat Provinsi Sumatera Barat Menurut Kabupaten/Kota Tahun 2021 berdasarkan validitas *cluster* karena memiliki nilai *Dunn Index* yang besar yaitu 1,22 dan nilai *Connectivity* yang kecil yaitu 5,85.

Kata Kunci: Kesejahteraan di Sumatera Barat, *Average Linkage*, *Ward*, Validitas *Cluster*

**PERFORMANCE COMPARISON OF AVERAGE LINKAGE METHODS  
AND WARD IN LEVELS GROUPING PROVINCIAL COMMUNITY  
WELFARE WEST SUMATERA BY DISTRICT /CITY YEAR 2021**

**Riska Ramadani**

**ABSTRACT**

Welfare in West Sumatra Province is still an important problem that is being faced by the government and also the entire community in 2021. The impact of the Covid-19 pandemic that occurred in early 2020 still leaves complicated issues that have not been resolved until 2021. The government is still trying to improve performance to improve the welfare of the community from all aspects of life. It is necessary to solve the problem of community welfare because this problem can have an impact on other social problems, such as crime problems, government problems, and other problems. Therefore, it is necessary to classify the level of community welfare in West Sumatra Province by Regency/City in 2021 using the average linkage and ward methods. Average Linkage method groups objects based on their closest distance using the Euclidean distance and the calculation of the average distance. While the ward method groups objects based on the Sum of Squares Error (SSE) value.

This research is an applied research which begins by studying the relevant theories with the problem. The data used is secondary data obtained from the Website Central Statistics Agency (BPS) of West Sumatra. The purpose of this research is to classify districts/cities based on the factors that affect the welfare of the people in the province of West Sumatra.

The results showed that using the average linkage and ward methods showed that the clusters formed were the same. It consists of 3 clusters, namely cluster 1 with a low level of welfare, cluster 2 with a moderate level of welfare and cluster 3 with a high level of welfare. Of the two methods, the average linkage method is better compared to the ward method in classifying the level of welfare of the people of West Sumatra Province by Regency/City in 2021 based on cluster validity because it has a large Dunn Index value is 1,22 and a small Connectivity value 5,85.

Keywords: Welfare in West Sumatra, Average Linkage, Ward, Cluster Validity

## KATA PENGANTAR



*Bismillahirrahmanirrahiim, Alhamdulillahirrabil'alamiin*, segala puji dan syukur penulis ucapkan kepada Allah SWT yang telah memberikan rahmat iman, islam, ikhsan. Serta nikmat pikiran sehingga penulis dapat menimba ilmu kepada guru-guru yang memiliki keahlian di bidang Statistika. Dari ilmu yang telah diberikan, penulis dapat menyelesaikan Skripsi dengan judul “Perbandingan Kinerja Metode *Average Linkage* dan *Ward* dalam Pengelompokan Tingkat Kesejahteraan Masyarakat Provinsi Sumatera Barat Menurut Kabupaten/Kota Tahun 2021”. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan Program Studi Statistika, Jurusan Statistika, Fakultas Matematika dan ilmu Pengetahuan Alam, Universitas Negeri Padang. Dalam menyelesaikan Skripsi ini penulis mendapatkan banyak bantuan dan dukungan dari berbagai pihak. Untuk itu penulis ucapkan terimakasih kepada:

1. Bapak Dr. Dony Permana, M.Si. Ketua Jurusan Statistika sekaligus Ketua Program Studi S1 Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang yang telah memberi arahan selama menuntut ilmu.
2. Ibu Dra. Nonong Amalita, M.Si. Sekretaris Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang yang telah memberi arahan dan dukungan selama menuntut ilmu.

3. Ibu Admi Salma, S.Pd., M.Si. Dosen Pembimbing Skripsi sekaligus Dosen Pembimbing Akademik yang telah memberikan arahan, bimbingan, dukungan dan motivasi dari awal sampai proses penyusunan Skripsi.
4. Bapak Dr. Syafriandi, M.Si. dan Bapak Dodi Vionanda, M.Si., Ph.D. Dosen Penguji Skripsi yang telah memberikan kontribusi terhadap Skripsi ini.
5. Bapak dan Ibu Dosen serta Staf Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang yang telah membantu penulis selama menimba ilmu di Program Studi Statistika.
6. Terkhusus Orangtua penulis, Bapak Adiswar dan Ibu Martanti yang telah berjuang, mendukung secara materil dan non materil serta mendo'akan penulis tanpa henti dalam proses menimba ilmu pada Program Studi Statistika, Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.
7. Rekan-rekan Program Studi S1 Statistika yang telah berjuang bersama selama perkuliahan.

Semoga Skripsi ini memberikan manfaat untuk penulis sendiri, bermanfaat untuk semua pihak, dan bernilai ibadah di sisi Allah SWT. Skripsi ini tidak terlepas dari kesalahan dan kekeliruan, oleh sebab itu penulis mengharapkan saran dan kritik yang bersifat membangun. Akhir kata penulis ucapkan terima kasih.

Padang, 10 Februari 2022

Riska Ramadani

## DAFTAR ISI

<b>ABSTRAK</b> .....	i
<b>KATA PENGANTAR</b> .....	iii
<b>DAFTAR ISI</b> .....	v
<b>DAFTAR GAMBAR</b> .....	vii
<b>DAFTAR TABEL</b> .....	viii
<b>DAFTAR LAMPIRAN</b> .....	ix
<b>BAB I PENDAHULUAN</b> .....	1
A. Latar Belakang Masalah.....	1
B. Batasan Masalah Penelitian.....	7
C. Rumusan Masalah Penelitian .....	8
D. Tujuan Penelitian .....	8
E. Manfaat Penelitian .....	9
<b>BAB II KERANGKA TEORITIS</b> .....	10
A. Kajian Teori .....	10
B. Penelitian yang Relevan .....	32
<b>BAB III METODOLOGI PENELITIAN</b> .....	34
A. Jenis Penelitian.....	34
B. Jenis dan Sumber Data Penelitian .....	34
C. Variabel Penelitian .....	34
D. Teknik Analisis Data .....	36
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	41
A. Hasil.....	41
B. Pembahasan.....	65
<b>BAB V PENUTUP</b> .....	69
A. Kesimpulan .....	69
B. Saran .....	70
<b>DAFTAR PUSTAKA</b> .....	71
<b>LAMPIRAN</b> .....	73

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
1. Pembagian Metode dalam Analisis Cluster .....	13
2. Single Linkage .....	17
3. Complete Linkage .....	18
4. Average Linkage .....	19
5. Ward.....	24
6. Contoh Dendogram .....	25
7. Diagram Alir Metode Average Linkage .....	37
8. Diagram Alir Metode Ward .....	39
9. Diagram Batang Jumlah Penduduk Menurut Kabupaten/Kota .....	42
10. AHH Menurut Kabupaten/Kota .....	42
11. HLS Menurut Kabupaten/Kota .....	43
12. Jumlah Penduduk Miskin menurut Kabupaten/Kota di Sumatera Barat Tahun 2021.....	44
13. PDRB Menurut Kabupaten/Kota .....	44
14. TPT Menurut Kabupaten/Kota.....	45
15. Penerangan Listrik Menurut Kabupaten/Kota .....	46
16. Sumber Air Minum Bersih Menurut Kabupaten/Kota.....	46
17. AMH Menurut Kabupaten/Kota .....	47
18. Jumlah Puskesmas Menurut Kabupaten/Kota.....	48
19. Dendogram Average Linkage .....	52
20. Dendogram Average Linkage dengan Partisi .....	53
21. Dendogram Ward .....	60
22. Dendogram Ward dengan Partisi .....	62

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1. Kesejahteraan Provinsi Sumatera Barat Tahun 2020 dan 2021 Menurut Faktor Utama yang Mempengaruhi .....	5
2. Perbedaan Teknik Dependen dan Interdependen .....	10
3. Variabel yang Digunakan dalam Penelitian .....	35
4. Struktur Data Penelitian .....	36
5. Anggota Cluster Average Linkage .....	53
6. Interpretasi Cluster Metode Average Linkage .....	54
7. Anggota Cluster Ward.....	62
8. Interpretasi Cluster Metode Average Linkage dan Ward .....	63
9. Validasi Cluster Metode Average Linkage dan Ward .....	65

## DAFTAR LAMPIRAN

<b>Lampiran</b>	<b>Halaman</b>
1. Data Penelitian .....	73
2. Standardisasi Data.....	74
3. Matriks Jarak Euclidean.....	75
4. Matriks Jarak Squared Euclidean.....	77
5. Nilai SSE Antar Objek.....	79

# BAB I

## PENDAHULUAN

### A. Latar Belakang Masalah

Analisis multivariat merupakan analisis yang dapat digunakan untuk melakukan analisis terhadap lebih dari dua variabel secara bersamaan. Analisis multivariat terbagi dua yaitu teknik dependen dan teknik interdependen. Pada teknik dependen terdapat dua jenis variabel yaitu variabel respon dan variabel penjelas. Teknik ini digunakan untuk menyelesaikan masalah mengenai hubungan antara dua jenis variabel yaitu variabel respon dan variabel penjelas. Sedangkan 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 (Hair *et al*, 2013: 5)

Menurut Rencher & Christensen (2012:501-539) salah satu analisis multivariat teknik interdependen adalah analisis *cluster*. Analisis *cluster* adalah pengelompokan objek dari suatu gugus data pengamatan kedalam suatu kelompok, dimana objek yang memiliki karakteristik sama akan digabungkan kedalam satu *cluster* yang sama, 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). Bagian *agglomerative* terdiri dari metode *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward*. Sedangkan salah satu metode non hirarki yaitu metode *K-Means*.

Metode *cluster* hirarki *agglomerative* dipilih karena metode ini lebih sering digunakan dalam proses pengelompokan. Kompleks dalam proses perhitungan pengelompokan objek karena memiliki berbagai metode yang dapat digunakan dalam pengelompokan sehingga metode tersebut dapat dibandingkan dalam memilih metode terbaik. Selain itu, metode ini mengelompokkan objek menjadi satu *cluster* berdasarkan karakteristik yang dimiliki oleh masing-masing objek. Metode *divisive* yang memecah satu *cluster* besar menjadi beberapa *cluster* kecil serta harus membentuk *cluster* berdasarkan perbedaan karakteristik yang dimiliki objek dan metode ini lebih sesuai digunakan untuk data yang memiliki nilai amatan yang hilang (Everitt *et al*, 2011:73-85). Dalam praktik yang dilakukan, kesulitan dalam melakukan pemecahan pada *cluster* yang besar pada metode *divisive* membuat metode *agglomerative* lebih banyak digunakan dalam pengelompokan (King, 2015:35). Menurut Kassambara (2017:72) pada akhir analisis *cluster* akan membentuk dendogram (pohon), sehingga mempermudah penafsiran dan tidak perlu menentukan banyak *cluster* yang akan dibentuk diawal karena bisa dilihat dari dendogram tersebut dan disesuaikan dengan kebutuhan penelitian.

Metode *Average Linkage* menurut Johnson & Wichern (2014: 690) merupakan metode analisis *cluster* hirarki *agglomerative* dimana metode ini mengelompokkan objek berdasarkan jarak rata-rata yang didapat dengan melakukan rata-rata semua jarak antar objek terlebih dahulu. Jarak antara dua *cluster* didefinisikan sebagai rata-rata jarak antara semua pasangan objek, di mana salah satu anggota dari pasangan berasal dari setiap *cluster*. 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* (Everitt *et al*, 2011: 79).

Metode *Ward* menurut Johnson & Wichern (2014: 692-693) 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*. Kelebihan metode ini dapat membentuk *cluster* yang memiliki varians dalam *cluster* yang kecil sehingga antar objek dalam *cluster* memiliki homogenitas atau kesamaan yang tinggi (Punj & Stewart, 1983).

Untuk menentukan perbandingan 2 (dua) metode *cluster* dapat menggunakan nilai *Dunn Index (DI)* dan nilai *Connectivity*. Nilai *DI* dan *Connectivity* pada masing-masing metode akan dihitung kemudian dibandingkan. Nilai *DI* diharapkan besar dan nilai *Connectivity* diharapkan kecil untuk menunjukkan metode *cluster* yang terbaik (Brock *et al*, 2008).

Berdasarkan pemaparan metode *cluster* tersebut, maka penerapan yang akan dilakukan yaitu pada permasalahan kesejahteraan yang terjadi di Provinsi Sumatera Barat tahun 2021. Pengelompokan dilakukan untuk mengetahui tingkat kesejahteraan masyarakat menurut objek yaitu Kabupaten/Kota di Sumatera Barat. *Cluster* dibentuk dengan proses *average linkage* dan *Ward*. Kedua metode ini akan menampilkan masing-masing dendogram. Penentuan tingkat kesejahteraan masyarakat menurut objek yaitu Kabupaten/Kota dapat ditentukan berdasarkan dendogram tersebut. Serta pengelompokan dilakukan karena belum ditemukan sumber yang menjelaskan mengenai tingkat kesejahteraan menurut

Kabupaten/Kota di Sumatera Barat yang ditentukan berdasarkan faktor-faktor yang mempengaruhi kesejahteraan yang diperoleh dari variabel-variabel yang ada dalam indikator kesejahteraan. Hal ini akan berguna untuk menunjukkan tingkat kesejahteraan masing-masing Kabupaten/Kota dan berguna dalam mengambil kebijakan dalam mengatasi masalah kesejahteraan masyarakat.

Kesejahteraan menurut referensi Anggaran Pendapatan Belanja Negara (APBN) Publik yang dikeluarkan oleh Dewan Perwakilan Rakyat Republik Indonesia (DPR RI) adalah segala sesuatu yang berkaitan langsung dengan tujuan Negara Indonesia. 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 (Mulia & Saputra, 2020). Terwujudnya kesejahteraan yang baik dapat menciptakan kehidupan masyarakat yang seimbang dan teratur, memberi peluang kepada seluruh lapisan masyarakat untuk memperoleh kehidupan yang layak, aman, selamat, adil dan makmur.

Kesejahteraan di Provinsi Sumatera Barat masih menjadi masalah penting yang sedang 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 yang dapat dilihat dari beberapa faktor utama yang mempengaruhi seperti yang terlihat pada Tabel 1.

**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
Perumahan (Milik Sendiri)	71,24%	68,34%

Sumber: BPS Sumatera Barat 2021

Pada Tabel 1 diperoleh informasi bahwa jumlah penduduk Sumatera Barat mengalami peningkatan pada tahun 2021. Peningkatan jumlah penduduk ini akan berdampak negatif terhadap meningkatnya persentase keparahan kemiskinan dan meningkatnya jumlah penduduk miskin yang akan mempengaruhi kesejahteraan masyarakat. AHH yang mengalami peningkatan sebesar 0,12 poin bisa diartikan baik yang menunjukkan perbaikan kesehatan masyarakat kearah positif. Namun berarti buruk karena semakin banyak bayi yang lahir, jumlah penduduk usia produktif semakin banyak maka masalah kesejahteraan sosial baru bisa saja muncul dimasa mendatang. Perumahan juga dijadikan sebagai tolok ukur kesejahteraan, dimana tempat yang dijadikan berlindung layak huni dan milik sendiri. Pada tahun 2021 perumahan milik sendiri mengalami penurunan sebesar 2,9% dari tahun 2020.

Hal ini bermakna negatif karena menunjukkan bahwa kesejahteraan masyarakat dilihat dari segi perumahan mengalami penurunan.

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 (Izzati, 2021). Secara keseluruhan, Sumatera Barat belum mampu mengendalikan pandemi Covid-19. Kasus harian Covid-19 yang fluktuatif membuat pemerintah mengalami kesulitan dalam mengambil kebijakan. Seperti yang terjadi pada tahun 2021, pemerintah memberlakukan pembatasan kegiatan masyarakat (PPKM) di beberapa Kabupaten/Kota yang berada dalam zona merah. Hal ini mengganggu stabilitas roda perekonomian antar daerah di Sumatera Barat yang berakibat kepada menurunnya kesejahteraan masyarakat. Banyak dari masyarakat yang menjadi pengangguran karena harus berhenti bekerja, pemotongan upah/gaji, bahkan diberhentikan tanpa pesangon.

Dari persoalan-persoalan yang terjadi itulah jumlah penduduk miskin akan meningkat, taraf kehidupan serta harapan hidup masyarakat akan melemah. Kemiskinan yang terjadi akan menimbulkan berbagai persoalan lainnya ditengah masyarakat seperti kriminalitas yang meningkat. Dampak lain dari persoalan kesejahteraan adalah ketidakteraturan sistem birokrasi pemerintahan akibat hilangnya kepercayaan masyarakat kepada penguasa.

Berdasarkan pemaparan yang telah dibahas, maka penulis melakukan penelitian yang berjudul “**Perbandingan Kinerja Metode *Average Linkage* dan**

## **Ward dalam Pengelompokan Tingkat Kesejahteraan Masyarakat Provinsi Sumatera Barat Menurut Kabupaten/Kota Tahun 2021”.**

### **B. Batasan Masalah Penelitian**

Agar pembahasan dan materi penelitian tidak meluas dan tidak menimbulkan kesalahan serta multitafsir pada masalah yang dibahas sehingga perlu dilakukan pembatasan. Secara keseluruhan banyak faktor-faktor yang dapat mempengaruhi kesejahteraan masyarakat di suatu daerah. Menurut BPS ada sekitar 29 faktor-faktor yang mempengaruhi kesejahteraan di Provinsi Sumatera Barat.

Oleh karena itu, batasan masalah pada penelitian ini meliputi variabel-variabel yang mempengaruhi kesejahteraan masyarakat Sumatera Barat tahun 2021 yang digunakan yaitu jumlah penduduk menurut kabupaten/kota, Angka Harapan Hidup menurut kabupaten/kota, Harapan Lama Sekolah menurut kabupaten/kota, jumlah penduduk miskin menurut kabupaten/kota, Pendapatan Domestik Regional Bruto (PDRB) menurut kabupaten/kota, Tingkat Pengangguran Terbuka menurut kabupaten/kota, persentase rumah tangga yang memiliki penerangan menurut kabupaten/kota, persentase rumah tangga menurut sumber air minum bersih menurut kabupaten/kota, Persentase Angka Melek Huruf Menurut Kabupaten/Kota, jumlah puskesmas menurut kabupaten/kota. Jadi, variabel yang digunakan dalam penelitian ini berjumlah sepuluh (10) variabel.

Metode yang digunakan untuk mengelompokkan tingkat kesejahteraan di Sumatera Barat menurut Kabupaten/Kota dengan analisis *cluster* metode *Average Linkage* dan metode *Ward*.

### **C. Rumusan Masalah Penelitian**

Berdasarkan latar belakang masalah yang telah dibahas, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana hasil pengelompokkan 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 pengelompokkan tingkat kesejahteraan masyarakat Provinsi Sumatera Barat menurut Kabupaten/Kota Tahun 2021?

### **D. Tujuan Penelitian**

Berdasarkan rumusan masalah penelitian, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui hasil pengelompokkan tingkat kesejahteraan masyarakat Provinsi Sumatera Barat menurut Kabupaten/Kota Tahun 2021 menggunakan metode *Average Linkage* dan *Ward*.
2. Untuk mengetahui metode terbaik dari metode *Average Linkage* dan *Ward* dalam pengelompokkan tingkat kesejahteraan masyarakat Provinsi Sumatera Barat menurut Kabupaten/Kota Tahun 2021.

### **E. Manfaat Penelitian**

Manfaat dari penelitian yang ingin dicapai adalah sebagai berikut:

1. Bagi peneliti, dapat menambah dan memperbaharui wawasan mengenai kesejahteraan khususnya di Provinsi Sumatera Barat.
2. Bagi masyarakat, diharapkan memberikan pengetahuan dan memberikan manfaat agar dapat selalu memperhatikan kesejahteraan. Sehingga dapat memperoleh kehidupan yang layak, nyaman, damai dan tentram.
3. Bagi pembuat kebijakan, diharapkan penelitian ini dapat dijadikan sebagai salah satu acuan untuk membuat keputusan yang bijaksana sehingga dapat mengatasi masalah kesejahteraan khususnya di Provinsi Sumatera Barat.
4. Bagi peneliti selanjutnya, diharapkan agar dapat meningkatkan, memperdalam serta memperluas kajian mengenai kesejahteraan dan metode yang digunakan. Karena topik ini sangat bermanfaat bagi kehidupan masyarakat dan metode yang digunakan merupakan metode yang berkembang sehingga akan fleksibel digunakan untuk penelitian selanjutnya.

## BAB II

### KERANGKA TEORITIS

#### A. Kajian Teori

##### 1. Analisis Multivariat

Menurut Hair *et al* (2013 : 4-5) analisis multivariat terbagi dua yaitu teknik dependen dan teknik interdependen. Pada teknik dependen terdapat dua jenis variabel yaitu variabel respon dan variabel penjelas. Teknik ini digunakan untuk menyelesaikan masalah mengenai hubungan antara dua jenis variabel yaitu variabel respon dan variabel penjelas. Sedangkan 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. Berikut ini tabel perbedaan teknik dependen dan interdependen pada Tabel 2.

**Tabel 2. Perbedaan Teknik Dependen dan Interdependen**

Perbedaan	Dependen	Interdependen
Jenis Variabel	Respon dan Penjelas	Penjelas
Jumlah variabel	1. Respon : 1, 2, atau lebih 2. Penjelas : memiliki lebih dari 2 variabel penjelas	Memiliki lebih dari dua variabel penjelas
Metode analisis data	Regresi linear berganda, regresi logistik, korelasi, dan lainnya	Analisis faktor, analisis <i>cluster</i>
Jenis variabel	Data numerik dan Kategorik	Data numerik

Analisis multivariat menurut Johnson & Wichern (2014: 1-5) adalah analisis yang digunakan untuk menganalisis banyak variabel secara bersamaan. Analisis ini diperlukan untuk mengetahui hubungan antara banyak variabel melalui berbagai teknik analisis multivariat. Metode analisis multivariat dapat digunakan dalam reduksi data yang berguna untuk melakukan analisis pada gugus data dengan menggunakan sedikit variabel tapi tidak membuat kehilangan informasi yang ada pada gugus data sehingga dalam merepresentasikan hasil yang diperoleh lebih mudah. Metode lain dari analisis multivariat dapat digunakan dalam pengelompokan berdasarkan kemiripan objek berdasarkan karakteristik yang dimiliki oleh objek tersebut. Karakteristik ini diperoleh dari variabel-variabel yang ada pada gugus data yang akan dikelompokkan. Serta analisis multivariat dapat digunakan dalam menganalisis hubungan sebab akibat antar variabel biasa disebut dengan tekni dependen.

Data multivariat digunakan untuk menyelidiki fenomena sosial atau fenomena fisik, dimana data multivariat terdiri dari banyak variabel. Setiap variabel memiliki karakter yang dicatat. Nilai dari variabel ini dicatat berdasarkan setiap objek yang berbeda, individu atau unit penelitian. Matriks data dapat ditulis sebagai berikut.

$$X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \text{ atau}$$

$X$  bisa dinyatakan sebagai  $X = \{x_{ik}\}$ , untuk  $i = 1, 2, \dots, n$  dan  $k = 1, 2, \dots, p$

$$\mathbf{x}_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} \text{ atau}$$

$$X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$$

Untuk mencari rata-rata  $n$  buah sampel pada variabel ke- $k$  adalah

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}$$

Keterangan:

$X$  : Matriks data

$n$  : Banyak amatan

$p$  : Banyak variabel

$\mathbf{x}_1$  : Vektor kolom untuk variabel pertama

$\bar{x}_k$  : Rata-rata variabel ke- $k$  untuk  $k = 1, 2, \dots, p$

## 2. Analisis Cluster

### a. Pengertian Analisis Cluster

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 (Rencher & Christensen, 2012:501).

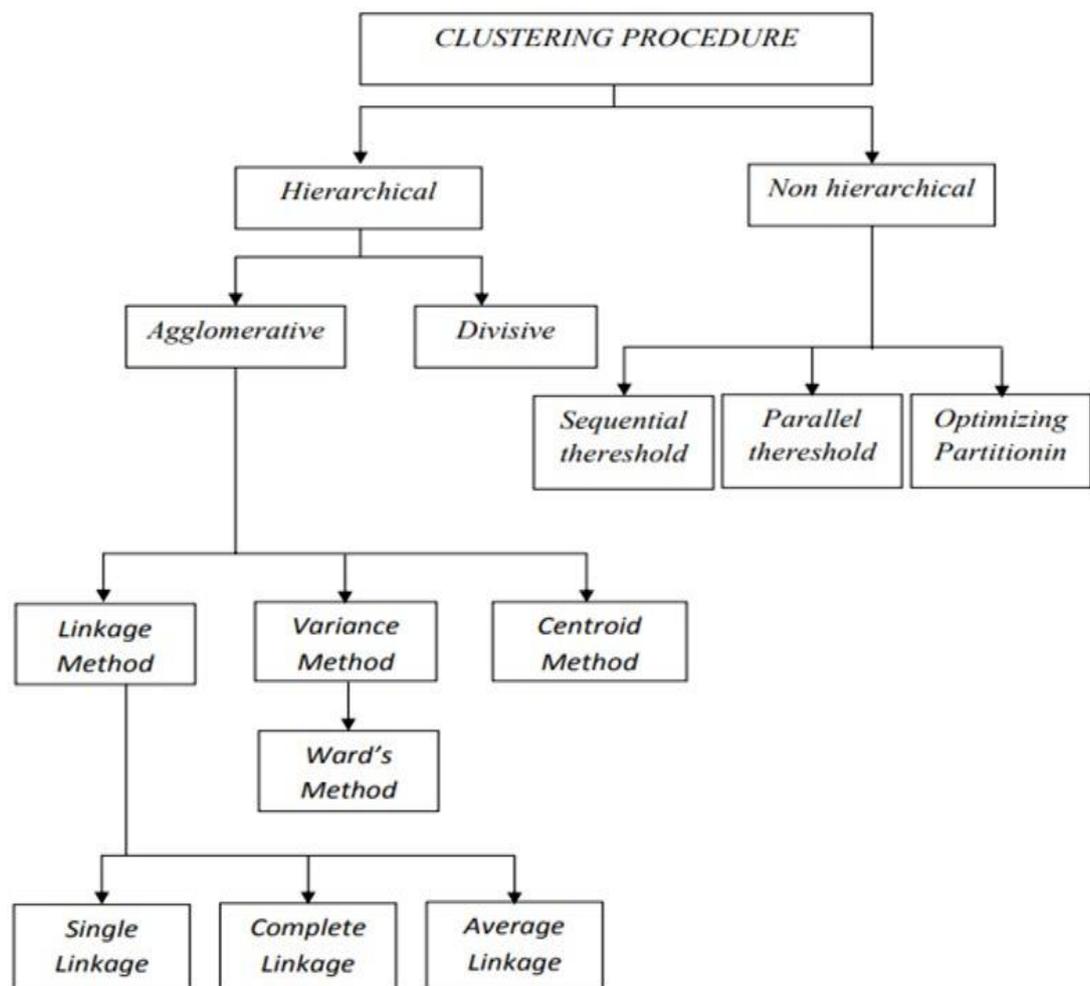
Jadi, analisis *cluster* adalah analisis multivariat yang digunakan untuk mengelompokkan objek yang memiliki kesamaan karakteristik kedalam satu

*cluster* dan objek yang memiliki karakteristik yang berbeda dimasukkan kedalam *cluster* lainnya, sehingga antar *cluster* memiliki karakteristik yang tidak sama.

Menurut Hair (2013:420) ciri-ciri *cluster* yang baik dari analisis *cluster* adalah :

- 1) Homogenitas atau kesamaan yang tinggi antar objek dalam satu *cluster* (*Within-Cluster*).
- 2) Heterogenitas atau perbedaan yang tinggi antar *cluster* yang satu dengan *cluster* yang lain (*Between-Cluster*).

Metode pada analisis *cluster* dapat dilihat pada Gambar 1.



Sumber: Simamora, 2005:215

**Gambar 1. Pembagian Metode dalam Analisis Cluster**

## b. Ukuran Jarak Analisis *Cluster*

Menurut Rencher & Christensen (2012:502-504), untuk menggunakan analisis *cluster* dapat diidentifikasi dari objek yang sama dan mengelompokkan kedalam satu *cluster*, banyak teknik yang dapat mengukur kesamaan atau kedekatan antara setiap pasang dari objek. Pengukuran kedekatan jarak yang tepat antara dua objek. Kemudian, objek yang memiliki jarak yang jauh termasuk kedalam pengukuran dari ketidaksamaan.

Pengukuran jarak yang biasa digunakan dalam analisis *cluster* adalah jarak *Euclidean* dengan rumus sebagai berikut.

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

Ada juga jarak *Squared Euclidean* yang merupakan variasi dari jarak *Euclidean* dengan rumus sebagai berikut :

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

Keterangan:

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

$p$  : banyak peubah yang digunakan

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

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

Penggunaan jarak *Euclidean* dalam proses pengelompokan dalam analisis *cluster* digunakan untuk data yang distandardisasi (Kassambara, 2017: 29).

### c. Standardisasi Data

Menurut Kassambara (2017: 28) standardisasi data dilakukan pada data yang memiliki perbedaan nilai yang cukup besar antar variabel. Perbedaan nilai yang berbeda akan menyebabkan perhitungan pada analisis *cluster* tidak benar-benar valid. Oleh sebab itu perlu dilakukan proses standardisasi data melalui transformasi pada data aktual kedalam bentuk *z-skor* dengan rumus sebagai berikut :

$$z_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k} \quad (3)$$

Keterangan:

$z_{ik}$  : nilai *z* baris ke-*i* kolom ke-*k*

$x_{ik}$  : nilai *x* baris ke-*i* kolom ke-*k*

$\bar{x}_k$  : rata-rata variabel kolom ke-*k*

$s_k$  : simpangan baku kolom ke-*k*

### d. Analisis Cluster Hierarki

Menurut Kassambara (2017:65) analisis *cluster* terbagi menjadi 2 bagian yaitu hierarki dan non hirarki. Metode *cluster* hierarki adalah pendekatan alternatif yang digunakan untuk mengelompokkan objek berdasarkan kesamaannya. Pada metode hirarki tidak memerlukan terlebih dahulu banyaknya kelompok atau *cluster* yang akan dihasilkan. Metode hirarki terbagi menjadi dua (2) tipe yaitu *agglomerative* (penggabungan) dan *divisive* (pemecahan). Tipe *agglomerative* yaitu teknik pengelompokan, dimana setiap objek atau pengamatan dianggap sebagai satu *cluster* sendiri, kemudian objek dalam masing-masing *cluster* yang memiliki kemiripan atau kesamaan digabungkan sampai menjadi satu *cluster* yang

besar. Sedangkan tipe *divisive* yaitu teknik pengelompokan, dimana semua objek berada dalam satu *cluster* besar, kemudian *cluster* yang memiliki karakteristik yang berbeda membentuk *cluster* yang lain sampai semua objek atau pengamatan berada *cluster* nya masing-masing.

Analisis *cluster* bagian hierarki dengan metode *agglomerative* dibagi kedalam beberapa metode sebagai berikut.

1) *Single Linkage* atau Pautan Tunggal

Menurut Johnson & Wichern (2014: 682) metode pautan tunggal dilakukan dengan meminimumkan jarak antar *cluster* yang digabungkan, atau dengan kata lain metode yang mengelompokkan dua objek yang mempunyai jarak terdekat terlebih dahulu. Jarak antar *cluster* di bentuk dari individu-individu dalam dua *klaster* yang mempunyai jarak terkecil atau kemiripan terbesar. Metode ini dimulai dengan menentukan jarak terkecil dalam matriks *proximity*  $D = \{d_{ik}\}$  dan gabungan objek yang memiliki jarak terkecil tersebut, misal U dan V, sehingga diperoleh *cluster* UV. Untuk mencari *cluster* UV dan W lainnya dapat diperoleh dengan cara sebagai berikut:

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\}$$

Dimana  $d_{UW}$  adalah jarak terdekat dari *cluster* U dan W sedangkan  $d_{VW}$  merupakan jarak terdekat dari *cluster* V dan W. Gambar 2 dari *Single Linkage*.



Sumber: Johnson & Wichern (2014: 681)

**Gambar 2. Single Linkage**

Gambar 2 menunjukkan jarak minimum pada metode pengelompokan *Single Linkage*. Maksud dari Gambar 2 adalah mengelompokkan objek berdasarkan jarak terdekat. Misalkan objek 1 dan 2 berada pada satu *cluster* yang sama karena memiliki kemiripan yang diperoleh dari jarak yang dekat. Pengelompokan berikutnya terjadi pada *cluster* yang telah terbentuk sebelumnya dengan objek 3, 4 dan 5. Dari proses tersebut ternyata objek yang memiliki kemiripan dengan *cluster* yang telah terbentuk adalah objek 4 berdasarkan jarak terdekat.

## 2) *Complete Linkage* atau Pautan Lengkap

Metode ini hampir sama dengan metode pautan tunggal hanya saja metode pautan lengkap menggunakan jarak terjauh. Metode pautan lengkap adalah metode yang mengklasterkan dua objek yang mempunyai jarak terjauh terlebih dahulu. *Cluster* dimulai dengan mencari jarak pada matriks *proximity*  $D = \{d_{ik}\}$  dan penggabungan antara  $U$  dan  $V$  untuk mendapatkan klaster pertama atau  $UV$ . Selanjutnya jarak antara  $UV$  dan setiap klaster dihitung dengan:

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\}$$

Dimana  $d_{UW}$  adalah jarak terjauh dari *cluster*  $U$  dan  $W$  sedangkan  $d_{VW}$  merupakan jarak terjauh dari *cluster*  $V$  dan  $W$  (Johnson & Wichern, 2014 : 685-686). Gambar 3 dari *complete linkage*.



Sumber: Johnson & Wichern (2014: 681)

**Gambar 3. Complete Linkage**

Gambar 3 menunjukkan jarak maksimum pada metode pengelompokan *Complete Linkage*. Maksud dari Gambar 3 adalah mengelompokkan objek berdasarkan jarak terjauh. Misalkan objek 1 dan 2 berada pada satu *cluster* yang sama karena memiliki kemiripan yang diperoleh dari jarak yang dekat. Pengelompokan berikutnya terjadi pada *cluster* yang telah terbentuk sebelumnya dengan objek 3, 4 dan 5. Dari proses tersebut ternyata objek yang memiliki kemiripan dengan *cluster* yang telah terbentuk adalah objek 5 berdasarkan jarak terjauh.

### 3) *Average Linkage*

Metode *Average Linkage* memperlakukan jarak antara dua *cluster* sebagai jarak rata-rata antara semua pasangan dari objek dimana salah satu anggota dari pasangan dimiliki oleh setiap *cluster*.

Metode ini dimulai dengan mencari matriks jarak  $D = \{d_{ik}\}$  untuk memperoleh objek-objek yang paling dekat (yang memiliki kemiripan) misalnya U dan V. Objek-objek ini digabungkan sehingga akan terbentuk *cluster* UV. Selanjutnya dihitung jarak antara *cluster* UV dengan *cluster* W, dengan menggunakan rumus sebagai berikut:

$$d_{(UV)W} = \frac{\sum_{i=1}^n \sum_{k=1}^n d_{ik}}{N_{(UV)}N_W} \quad (4)$$

Keterangan:

$d_{(UV)W}$ : merupakan jarak antara *cluster UV* dengan *cluster V*

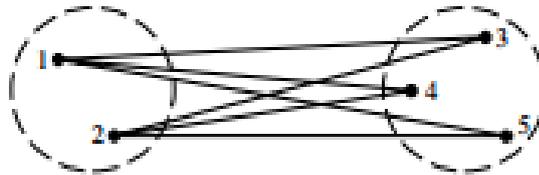
$d_{ik}$ : merupakan jarak antara objek  $i$  pada *cluster (UV)* dengan objek  $k$  pada *cluster*

$N_{(UV)}$ : merupakan jumlah objek pada *cluster (UV)*

$N_W$ : merupakan jumlah objek pada *cluster W*

(Johnson & Wichern, 2014 : 690)

Menurut Yim & Ramdeen (2015), metode *Average Linkage* digunakan ketika metode *Single Linkage* dan *Complete Linkage* memiliki keterbatasan saat menentukan ukuran jarak dalam menelompokkan karakteristik yang sama. Metode ini juga lebih stabil karena menghitung jarak berdasarkan rata-rata jarak dari dua objek. Gambar 4 *Average Linkage*.



Sumber: Johnson & Wichern (2014: 681)

**Gambar 4. Average Linkage**

Gambar 4 menunjukkan jarak rata-rata pada metode pengelompokan *Average Linkage*. Maksud dari Gambar 4 adalah mengelompokkan objek berdasarkan jarak rata-rata. Misalkan objek 1 dan 2 berada pada satu *cluster* yang sama karena memiliki kemiripan yang diperoleh dari jarak yang dekat. Pengelompokan berikutnya terjadi pada *cluster* yang telah terbentuk sebelumnya dengan objek 3, 4 dan 5 berdasarkan jarak rata-rata.

#### 4) Ward

Pada metode *Ward cluster* yang terbentuk memiliki varians internal sekecil mungkin karena jarak antara dua *cluster* didasarkan pada *Sum of Square Error*

(*SSE*) dari kedua *cluster* yang terbentuk. Metode ini juga efektif sehingga sering digunakan untuk dibandingkan dengan metode hirarki lainnya (Govender & Sivakumar, 2020).

*SSE* hanya dapat dihitung jika *cluster* memiliki elemen lebih dari satu objek. *SSE* cluster yang hanya memiliki satu objek adalah nol (0). Metode ini menggunakan perhitungan yang kompleks dan memaksimalkan homogenitas dalam satu cluster. Ukuran yang digunakan dalam metode ini adalah *SSE*. Secara umum, *SSE* dinotasikan sebagai berikut:

$$SSE = \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})' (\mathbf{x}_j - \bar{\mathbf{x}})$$

Keterangan:

$\mathbf{x}_j$  : vektor kolom yang berisi nilai objek  $j$  dengan  $j = 1, 2, \dots, n$  dalam *cluster*

$\bar{\mathbf{x}}$  : vektor kolom yang merupakan rata-rata nilai objek dalam *cluster*.

$n$  : banyaknya objek dalam *cluster* yang terbentuk.

Hasil dari metode *Ward* dapat ditampilkan dalam bentuk dendrogram. (Johnson & Wichern, 2014 : 692-693).

Misalkan  $U$  adalah *cluster*, maka jumlah kuadrat kesalahan dalam *cluster* tersebut adalah:

$$SSE_U = \sum_{i=1}^{n_U} (\mathbf{x}_i - \bar{\mathbf{x}}_U)' (\mathbf{x}_i - \bar{\mathbf{x}}_U)$$

Dan jika  $V$  adalah *cluster*, maka jumlah kuadrat kesalahan dalam *cluster* tersebut adalah:

$$SSE_V = \sum_{i=1}^{n_V} (\mathbf{x}_i - \bar{\mathbf{x}}_V)' (\mathbf{x}_i - \bar{\mathbf{x}}_V)$$

Jika  $UV$  adalah *cluster*, maka jumlah kuadrat kesalahan dalam *cluster* tersebut adalah:

$$SSE_{UV} = \sum_{i=1}^{n_{UV}} (\mathbf{x}_i - \bar{\mathbf{x}}_{UV})' (\mathbf{x}_i - \bar{\mathbf{x}}_{UV})$$

Keterangan:

$SSE_U, SSE_V, SSE_{UV}$  : jumlah kuadrat kesalahan pada *cluster*  $U, V$  dan  $UV$ .

$n_U, n_V, n_{UV}$  : banyak objek pada *cluster*  $U, V, UV$ .

$\bar{\mathbf{x}}_U, \bar{\mathbf{x}}_V, \bar{\mathbf{x}}_{UV}$  : vektor kolom yang entrinya rata-rata nilai objek dari *cluster*  $U, V$  dan  $UV$

$$\bar{\mathbf{x}}_U = \sum_{i=1}^{n_U} \frac{x_i}{n_U}$$

$$\bar{\mathbf{x}}_V = \sum_{i=1}^{n_V} \frac{x_i}{n_V}$$

$$\bar{\mathbf{x}}_{UV} = \frac{n_U \bar{\mathbf{x}}_U + n_V \bar{\mathbf{x}}_V}{n_U + n_V}$$

Metode *Ward* menggabungkan dua *cluster*  $U$  dan  $V$  dengan meminimumkan peningkatan  $SSE$  yang didefinisikan sebagai jarak antara *cluster*  $U$  dan *cluster*  $V$  sebagai berikut:

$$I_{UV} = SSE_{UV} - (SSE_U + SSE_V)$$

Hal ini dapat ditunjukkan bahwa peningkatan  $I_{UV}$  memiliki bentuk ekuivalen sebagai berikut:

$$\begin{aligned} I_{UV} &= n_U (\bar{\mathbf{x}}_U - \bar{\mathbf{x}}_{UV})' (\bar{\mathbf{x}}_U - \bar{\mathbf{x}}_{UV}) + n_V (\bar{\mathbf{x}}_V - \bar{\mathbf{x}}_{UV})' (\bar{\mathbf{x}}_V - \bar{\mathbf{x}}_{UV}) \\ &= n_U \left( \bar{\mathbf{x}}_U - \left( \frac{n_U \bar{\mathbf{x}}_U + n_V \bar{\mathbf{x}}_V}{n_U + n_V} \right) \right)' \left( \bar{\mathbf{x}}_U - \left( \frac{n_U \bar{\mathbf{x}}_U + n_V \bar{\mathbf{x}}_V}{n_U + n_V} \right) \right) + \\ &\quad n_V \left( \bar{\mathbf{x}}_V - \left( \frac{n_U \bar{\mathbf{x}}_U + n_V \bar{\mathbf{x}}_V}{n_U + n_V} \right) \right)' \left( \bar{\mathbf{x}}_V - \left( \frac{n_U \bar{\mathbf{x}}_U + n_V \bar{\mathbf{x}}_V}{n_U + n_V} \right) \right) \\ &= n_U \left( \frac{(n_U + n_V) \bar{\mathbf{x}}_U - n_U \bar{\mathbf{x}}_U - n_V \bar{\mathbf{x}}_V}{n_U + n_V} \right)' \left( \frac{(n_U + n_V) \bar{\mathbf{x}}_U - n_U \bar{\mathbf{x}}_U - n_V \bar{\mathbf{x}}_V}{n_U + n_V} \right) + \\ &\quad n_V \left( \frac{(n_U + n_V) \bar{\mathbf{x}}_V - n_U \bar{\mathbf{x}}_U - n_V \bar{\mathbf{x}}_V}{n_U + n_V} \right)' \left( \frac{(n_U + n_V) \bar{\mathbf{x}}_V - n_U \bar{\mathbf{x}}_U - n_V \bar{\mathbf{x}}_V}{n_U + n_V} \right) \end{aligned}$$

$$\begin{aligned}
&= n_U \left( \frac{n_U \bar{x}_U + n_V \bar{x}_U - n_U \bar{x}_U - n_V \bar{x}_V}{n_U + n_V} \right)' \left( \frac{n_U \bar{x}_U + n_V \bar{x}_U - n_U \bar{x}_U - n_V \bar{x}_V}{n_U + n_V} \right) + \\
&\quad n_V \left( \frac{n_U \bar{x}_V + n_V \bar{x}_V - n_U \bar{x}_U - n_V \bar{x}_V}{n_U + n_V} \right)' \left( \frac{n_U \bar{x}_V + n_V \bar{x}_V - n_U \bar{x}_U - n_V \bar{x}_V}{n_U + n_V} \right) \\
&= n_U \left( \frac{n_V \bar{x}_U - n_V \bar{x}_V}{n_U + n_V} \right)' + \left( \frac{n_V \bar{x}_U - n_V \bar{x}_V}{n_U + n_V} \right) + n_V \left( \frac{n_U \bar{x}_V - n_U \bar{x}_U}{n_U + n_V} \right)' + \left( \frac{n_U \bar{x}_V - n_U \bar{x}_U}{n_U + n_V} \right) \\
&= n_U \left( \frac{n_V (\bar{x}_U - \bar{x}_V)}{n_U + n_V} \right)' + \left( \frac{n_V (\bar{x}_U - \bar{x}_V)}{n_U + n_V} \right) + n_V \left( \frac{n_U (\bar{x}_U - \bar{x}_V)}{n_U + n_V} \right)' + \left( \frac{n_U (\bar{x}_U - \bar{x}_V)}{n_U + n_V} \right) \\
&= \frac{n_U n'_U n_V}{(n_U + n_V)' (n_U + n_V)} (\bar{x}_U - \bar{x}_V)' (\bar{x}_U - \bar{x}_V) + \frac{n_V n'_U n_U}{(n_U + n_V)' (n_U + n_V)} (\bar{x}_U - \\
&\quad \bar{x}_V)' (\bar{x}_U - \bar{x}_V) \\
&= \frac{n_U n'_U n_U n_V + n_V n'_U n_U n_V}{(n_U + n_V)' (n_U + n_V)} (\bar{x}_U - \bar{x}_V)' (\bar{x}_U - \bar{x}_V) \\
&= \frac{n_U (n_V n'_U) + n_V (n_U n'_U)}{(n_U + n_V)' (n_U + n_V)} (\bar{x}_U - \bar{x}_V)' (\bar{x}_U - \bar{x}_V) \\
&= \frac{n_U n_V (n'_U) + n_V n_U (n'_U)}{n_U n'_U + n_U n'_U + n_V n'_U + n_V n'_U} (\bar{x}_U - \bar{x}_V)' (\bar{x}_U - \bar{x}_V) \\
&= \frac{n_U n_V (n'_U + n'_U)}{n_U (n'_U + n'_U) + n_V (n'_U + n'_U)} (\bar{x}_U - \bar{x}_V)' (\bar{x}_U - \bar{x}_V) \\
&= \frac{n_U n_V (n'_U + n'_U)}{n_U + n_V (n'_U + n'_U)} (\bar{x}_U - \bar{x}_V)' (\bar{x}_U - \bar{x}_V) \\
&= \frac{n_U n_V}{n_U + n_V} ((\bar{x}_U - \bar{x}_V)' (\bar{x}_U - \bar{x}_V))
\end{aligned}$$

Meminimumkan peningkatan  $SSE$  ekuivalen dengan meminimumkan jarak antar objek. Jika  $U$  terdiri dari  $x_i$  dan  $V$  terdiri dari  $x_j$ , maka  $SSE_U$  dan  $SSE_V$  adalah nol (0). Selanjutnya diperoleh persamaan jarak antar objek yang merupakan rumus yang digunakan untuk menentukan jarak dengan metode *Ward* sebagai berikut:

$$I_{ik} = SSE_{ik} = \frac{1}{2} d_{ik}^2 \quad (5)$$

Keterangan:

$I_{ik}$  : merupakan jarak antara objek  $i$  dan objek  $k$

- $x_i$  : merupakan nilai rata-rata objek  $i$   
 $x_k$  : nilai rata-rata objek  $k$   
 $d_{ik}^2$  : jarak *Squared Euclidean* antara objek  $i$  dan objek  $j$   
 $k$  :  $1, 2, \dots, p$   
 $p$  : jumlah variabel *cluster*  
 $ik$  :  $1, 2, \dots, n$

Dari hasil jarak antar objek diperoleh *cluster* pertama yang memiliki jarak terdekat yang merupakan *cluster* yang paling homogen diantara *cluster* lain. Total jarak terdekat dihitung dengan rumus:

$$I = I_1 + I_2 + \dots + I_n$$

Keterangan:

- $I$  : Total jarak terdekat  
 $I_n$  : jarak terdekat ke- $n$

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

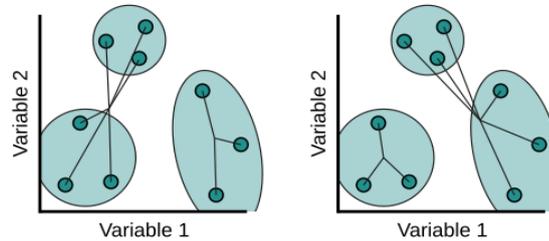
$$I_{(UV)W} = \frac{n_U + n_W}{n_{UV} + n_W} I_{UW} + \frac{n_V + n_W}{n_{UV} + n_W} I_{VW} - \frac{n_W}{n_{UV} + n_W} I_{UV} \quad (6)$$

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  $U, V, W$

(Rencher & Christensen, 2012:517-518)

Contoh metode *Ward* pada Gambar 5.

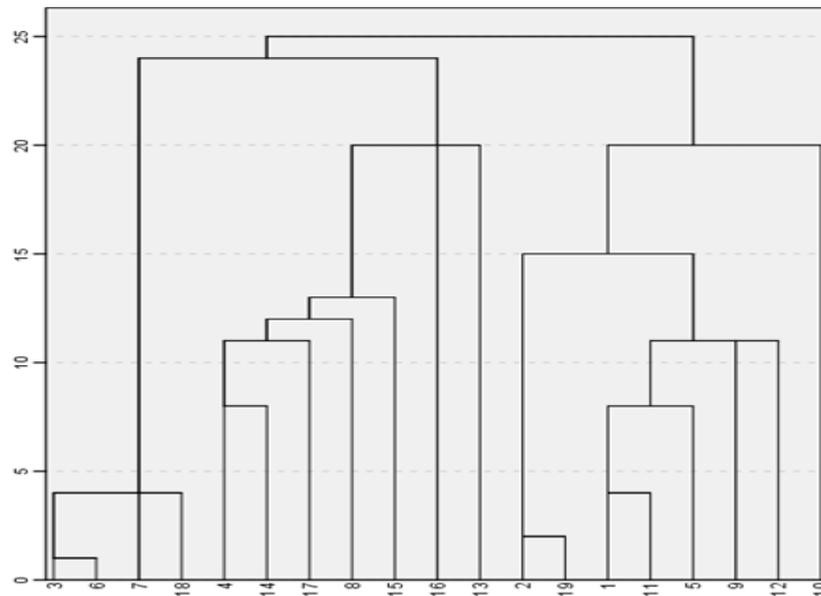


**Gambar 5. Ward**

Gambar 5 menunjukkan metode pengelompokan *Ward* dengan mencari nilai *SSE*. Dua objek akan bergabung menjadi satu *cluster* berdasarkan nilai *SSE* yang minimum. Kemudian pengelompokan berikut dengan *cluster* yang terbentuk dan objek-objek lain berdasarkan nilai *SSE* yang minimum.

#### e. Dendrogram atau Grafik Pohon

Menurut King (2015:12) hasil dari analisis cluster hirarki akan ditampilkan dalam bentuk dendrogram. Dendrogram merupakan diagram atau grafik pohon yang digunakan secara hirarki untuk mengelompokkan objek dalam analisis *cluster*. Dalam analisis *cluster* hirarki pengelompokan dimulai dengan dua objek yang mempunyai kesamaan paling dekat. Begitu seterusnya sampai ke objek yang lain, sehingga *cluster* tersebut akan membentuk semacam “pohon”. Dimana ada hirarki atau tingkatan yang jelas antar objek, dari yang paling mirip sampai yang tidak mirip. Contoh dari Gambar 6 mengenai dendrogram.



**Gambar 6. Contoh Dendrogram**

Pada Gambar 6 dapat dilihat contoh dendrogram hasil analisis *cluster* hirarki. Sumbu horizontal menunjukkan objek sedangkan sumbu vertikal menunjukkan ukuran jarak. Objek yang memiliki jarak terdekat akan bergabung menjadi satu *cluster*. Kemudian objek yang memiliki kemiripan dengan *cluster* yang telah terbentuk sebelumnya juga akan ikut bergabung menjadi satu *cluster*. Hal ini akan berlangsung secara terus-menerus sampai semua objek akan berada pada *cluster* masing-masing.

#### **f. Validasi Metode Analisis *Cluster* Hirarki**

Menurut Hennig *et al* (2016: 602-603) untuk mengetahui perbandingan dua metode yang memiliki cara kerja yang paling baik dapat ditentukan berdasarkan *Internal Measures*. *Internal Measures* adalah pengukuran untuk melakukan validasi *cluster* berdasarkan informasi internal yang terdapat pada data. Metode yang biasa digunakan yaitu *Dunn Index (DI)*. *DI* diperkenalkan oleh Dunn pada

tahun 1974. *DI* adalah ukuran validasi hasil cluster yang diperoleh dari mengukur jarak antara dua cluster dengan jarak maksimum dalam *cluster*. Nilai *DI* diharapkan tinggi untuk menunjukkan bahwa antar *cluster* memiliki perbedaan yang tinggi. *DI* dirumuskan sebagai berikut:

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

Keterangan:

*DI* : *Dunn Index*

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

$\text{diam}(c_m)$  : Diameter *cluster i* atau jarak dalam kelompok

Selanjutnya *internal measures* yang digunakan dalam analisis *cluster* adalah *Connectivity*. Menurut Kassambara (2017:139) *Connectivity* adalah nilai validitas yang berhubungan dengan mengavaluasi homogenitas dari cluster, biasa dilihat menggunakan varians *intra-cluster*. Hubungan ini menunjukkan posisi dari data objek dalam sebuah *cluster*, yang disebut tetangga terdekat. Nilai *Connectivity* berada antara nol sampai tak hingga. Nilai ini diharapkan sekecil mungkin untuk menunjukkan bahwa antar objek dalam *cluster* memiliki kesamaan yang tinggi.

*Connectivity* dirumuskan sebagai berikut:

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

*Conn* : *Connectivity*

$nn_{i(j)}$  : Objek terdekat dari data ke-*j* ke data *i*, jika berada dalam satu *cluster* bernilai 0 dan jika berbeda *cluster* bernilai  $1/j$

*N* : Banyak pengamatan

$L$  : Banyak *cluster*

**g. Interpretasi Analisis Cluster**

Interpretasi suatu *cluster* meliputi penggambaran karakteristik masing-masing *cluster* untuk 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:

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

Keterangan:

$v$  : Nilai *centroid* atau nilai rata-rata

$y_i$  : objek ke- $i$

$n$  : banyaknya objek

(Hair, 2015:450)

**3. R Packages Analisis Cluster Hirarki**

Untuk melakukan analisis *cluster* hirarki dengan menggunakan *software R* diperlukan *package*, yaitu *R Package cValid*. Pada *package* ini terdiri dari beberapa fungsi yang dapat digunakan dalam menentukan validasi analisis *cluster* hirarki. Fungsi yang digunakan dalam menentukan validasi *Dunn Index* dan *Connectivity* adalah *cValid* ( ). Pada fungsi ini terdiri dari beberapa bagian yaitu jarak yang digunakan, metode *cluster* hirarki yang digunakan, kelompok *cluster* yang terbentuk (Brock *et al*, 2021:7).

#### **4. Kesejahteraan**

Kesejahteraan menurut Sinaga (2016) merupakan suatu keadaan yang mengacu kepada kehidupan ekonomi dan sosial masyarakat. Kesejahteraan dapat diartikan sebagai tindakan yang dilakukan manusia dengan tujuan untuk memperoleh kehidupan masyarakat yang lebih baik. Sejalan dengan penjelasan sebelumnya, Perserikatan Bangsa-Bangsa (PBB) mengartikan kesejahteraan adalah keadaan yang terorganisir yang bertujuan untuk membantu penyesuaian hubungan mutualisme atau hubungan timbal balik yang menguntungkan antara individu dengan lingkungan.

Dengan demikian, kesejahteraan dapat disimpulkan sebagai bentuk keadaan kehidupan masyarakat yang berkaitan dengan lingkungan dengan harapan memperoleh kehidupan yang lebih baik. Tujuan yang dilakukan agar kesejahteraan yang baik dicapai melalui teknik dan metode yang telah ditetapkan dengan maksud agar individu, kelompok atau komunitas dapat menyelesaikan masalah yang berkaitan dengan perubahan pola kehidupan masyarakat, serta melakukan kerja sama untuk memperbaiki kondisi yang berkaitan dengan kesejahteraan yang meliputi perekonomian dan sosial.

Untuk menentukan tingkat kesejahteraan suatu masyarakat di suatu daerah, dapat dilihat melalui indikator dengan variabel yang ada dalam indikator tersebut.

#### **5. Kesejahteraan Masyarakat Provinsi Sumatera Barat dan Faktor yang Mempengaruhi**

Kesejahteraan masyarakat Sumatera Barat tahun 2021 dilihat dari beberapa faktor yang mempengaruhi sebagai berikut.

a. Jumlah Penduduk

Jumlah penduduk Sumatera Barat terus mengalami penambahan. Dibandingkan dengan hasil sensus penduduk tahun 2010 jumlah penduduk Sumatera Barat telah bertambah sekitar 633,42 ribu orang atau meningkat 13,02 persen selama sembilan tahun terakhir. Salah satu persoalan terkait dengan masalah jumlah penduduk adalah ketimpangan distribusi penduduk antar daerah kabupaten dan kota dan antara penduduk perdesaan dan perkotaan. Distribusi penduduk yang tidak merata bisa menimbulkan permasalahan disuatu daerah. Beberapa daerah kota penduduknya semakin besar sementara itu beberapa daerah kabupaten masih dihuni oleh penduduk yang relatif sedikit. Kondisi ini akan berpengaruh terhadap perkembangan masyarakat serta munculnya masalah ketidakmerataan kesejahteraan masyarakat (BPS Sumatera Barat, 2021: 3-4)

Laju pertumbuhan penduduk yang tidak terkendali tidak diimbangi dengan ketersediaan fasilitas dasar dan lapangan pekerjaan akan menimbulkan banyak permasalahan, seperti tingkat kemiskinan, pengangguran, kriminalitas, dan lain sebagainya.

b. Angka Harapan Hidup (AHH)

Menurut BPS (2020) AHH adalah rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang selama hidup. AHH dijadikan sebagai alat untuk mengevaluasi kinerja pemerintah dalam upaya peningkatan kesejahteraan masyarakat. AHH yang tinggi menunjukkan keberhasilan pembangunan, namun sebaliknya AHH yang rendah menunjukkan bahwa pembangunan dan pemerataan kesejahteraan masyarakat belum memuaskan.

c. Harapan Lama Sekolah (HLS)

HLS adalah lamanya sekolah dalam tahun yang diharapkan akan dirasakan oleh anak pada umur tertentu. HLS dijadikan sebagai faktor penentu dalam kesejahteraan masyarakat dalam bidang pendidikan (BPS, 2020).

d. Angka Melek Huruf (AMH)

Menurut BPS (2020) AMH adalah persentase penduduk berusia 15 tahun keatas yang dapat membaca dan menulis huruf latin dan huruf lainnya. AMH akan berpengaruh kepada kesejahteraan masyarakat karena AMH dapat menunjukkan suatu daerah sudah sejahtera dari segi pendidikan atau belum. Faktor ini menjadi penentu kesejahteraan masyarakat guna mencapai pembangunan daerah yang lebih baik.

e. Jumlah Penduduk Miskin

Kemiskinan yang merupakan masalah multidimensi yang memerlukan kebijakan dan program intervensi yang bersifat multidimensi pula agar kesejahteraan meningkat sehingga dapat terbebas dari masalah kemiskinan (Muda *et al*, 2019). Jumlah penduduk miskin di Sumatera Barat mengalami peningkatan pada tahun 2021 dibandingkan tahun 2020. Hal ini berakibat kepada salah satu faktor penghambat dalam meningkatnya kesejahteraan masyarakat. Hal ini juga berakibat kepada ketertinggalan pembangunan daerah.

f. Pendapatan Domestik Regional Bruto (PDRB)

Menurut BPS (2020) PDRB adalah jumlah keseluruhan dari nilai tambah barang dan jasa yang dihasilkan berdasarkan semua kegiatan perekonomian diseluruh wilayah dalam periode tahun tertentu yang pada umumnya dalam kurun

waktu satu tahun. PDRB dapat digunakan sebagai alat untuk menentukan kesejahteraan disuatu daerah. Jika PDRB tinggi maka dapat dikatakan masyarakat dalam daerah tersebut dalam keadaan sejahtera. PDRB juga berkaitan dengan Sumber Daya Manusia dan Sumber Daya Alam yang ada pada daerah tersebut.

g. Tingkat Pengangguran Terbuka

Menurut BPS (2020) TPT adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja. Persentase TPT yang rendah akan menunjukkan bahwa jumlah pengangguran suatu daerah sudah berkurang dan kesejahteraan telah meningkat. Namun, apabila TPT memiliki persentase yang tinggi maka jumlah pengangguran suatu daerah bertambah jumlahnya dan kesejahteraan masyarakatnya mengalami penurunan.

h. Persentase Rumah Tangga yang Memiliki Penerangan

Penentu kesejahteraan masyarakat dapat juga dilihat dari rumah tangga yang sudah memiliki akses dan jangkauan penerangan listrik. Hal ini didasarkan pada kemampuan masyarakat dalam membayar listrik serta daerah yang didiami sudah termasuk kedalam daerah yang mudah diakses.

i. Persentase Rumah Tangga Menurut Sumber Air Minum Bersih

Air minum yang bersih merupakan faktor kesejahteraan masyarakat paling penting karena dapat mewujudkan masyarakat yang sehat dan sejahtera. Air sangat berguna bagi kehidupan masyarakat, semua kegiatan dan aktivitas masyarakat sangat tergantung pada ketersediaan air terutama air yang bersih.

j. Jumlah Puskesmas

Adanya puskesmas dapat menjamin kehidupan yang sehat dan meningkatkan kesejahteraan seluruh masyarakat. Tersedianya puskesmas di setiap daerah baik daerah perkotaan maupun daerah terpencil sangat membantu masyarakat dalam memperoleh pengobatan yang lebih baik.

**B. Penelitian yang Relevan**

Beberapa penelitian yang telah menggunakan analisis *cluster* metode *Average Linkage* dan *Ward* antara lain.

1. Penelitian yang dilakukan oleh Silvia Ningsih, Sri Wahyu Ningsih dan Yuki Novia Nasution (2016) dengan judul Perbandingan Kinerja Metode *Complete Linkage* dan *Average Linkage* dalam Menentukan Hasil Analisis *Cluster* (Studi Kasus: Produksi Palawaija Provinsi Kalimantan Timur 2014/2015). Hasil penelitian yang diperoleh adalah metode *Average Linkage* lebih baik karena nilai rasio simpangan baku *cluster* lebih kecil yaitu sebesar 0.056 dari pada rasio simpangan baku *cluster* metode *Complete Linkage*.
2. Penelitian yang dilakukan oleh Nurhan Dogan, Ismet Dogan dan Sinan Saracli (2013) dengan judul *Comparison of Hierarchical Cluster Analysis Methods by Cophenetic Correlation* menunjukkan bahwa metode *Average Linkage* lebih baik dibandingkan metode *Single Linkage* dan *Complete Linkage*.
3. Penelitian yang dilakukan oleh Nurafiza Thamrin dan Arie Wahyu Wijayanto (2021) dengan judul Analisis *Cluster* dengan Menggunakan *Hard Clustering* dan *Soft Clustering* untuk Pengelompokan Tingkat Kesejahteraan

Kabupaten/Kota di Pulau Jawa menunjukkan bahwa metode *Ward* lebih baik dalam mengelompokan wilayah berdasarkan tingkat kesejahteraan dibandingkan metode *Single Linkage*.

4. Penelitian yang dilakukan oleh Vijaya, Shweta Sharma, Neha Batra (2019) dengan judul *Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering* metode *Ward* lebih baik dalam pengelompokan dibanding metode *single* dan *Complete Linkage*.

## BAB V

### PENUTUP

#### A. Kesimpulan

Berdasarkan hasil dan pembahasan yang didapatkan, maka dapat disimpulkan bahwa.

1. 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. Jumlah cluster adalah 3 dengan kategori tingkatan *cluster* yang terbentuk adalah rendah, sedang dan tinggi. *Cluster* 1 dengan kesejahteraan rendah dengan 1 objek yaitu Kabupaten Kepulauan Mentawai. *Cluster* 2 dengan kesejahteraan sedang dengan anggota 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. Dan *cluster* 3 dengan kesejahteraan tinggi anggota *cluster* ini yaitu Kota Padang.
2. 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*. Nilai *DI Average Linkage* lebih besar yaitu sebesar 1,22 dan nilai *Connectivity* yang lebih kecil yaitu sebesar 5,85 dibandingkan metode *Ward* dengan nilai *DI* sebesar 0,28 dan nilai *Connectivity* 7,92.

## **B. Saran**

Berdasarkan hasil, pembahasan serta kesimpulan yang diperoleh. Maka saran yang dapat diberikan adalah sebagai berikut.

1. Pada penelitian yang dilakukan dan ditulis pada skripsi ini membahas mengenai pengelompokan tingkat kesejahteraan masyarakat Provinsi Sumatera Barat dengan metode *Average Linkage* dan *Ward*. Oleh karena itu, peneliti selanjutnya yang juga akan menggunakan analisis *cluster* dalam pengelompokan berbagai permasalahan dapat menggunakan metode *cluster* yang lainnya. Karena analisis *cluster* merupakan metode yang berkembang sehingga akan cocok jika digunakan dalam segala jenis penelitian.
2. Selain itu, hasil penelitian yang sudah dilakukan sekiranya dapat dijadikan sebagai salah satu acuan dalam mengambil kebijakan supaya dapat meningkatkan kesejahteraan masyarakat terkhusus di Provinsi Sumatera Barat. Kebijakan yang sesuai dapat diberikan kepada Kabupaten Kepulauan Mentawai karena memiliki tingkat kesejahteraan paling rendah. Misalnya memanfaatkan Sumber Daya Alam yang tersedia di daerah tersebut.

## DAFTAR PUSTAKA

- Badan Pusat Statistik (BPS). “Indikator Kesejahteraan Rakyat Provinsi Sumatera Barat 2020”. Publikasi BPS diakses dari <http://www.sumbar.bps.go.id/> ,tanggal 01 Januari 2022.
- Badan Pusat Statistik (BPS). “Indikator Kesejahteraan Rakyat Provinsi Sumatera Barat 2021”. Publikasi BPS diakses dari <http://www.sumbar.bps.go.id/> ,tanggal 01 Januari 2022.
- Simamora, B. (2005). Analisis Multivariat Pemasaran. Jakarta: Gramedia.
- Brock,G., Datta, S., Datta, S.,& Pihur,V. (2008). cIVAlid: An R Package for Cluster Validation. *Journal of Statistical Software*, 25(4), 1-22. <http://www.jstatsoft.org/>
- Brock,G., Datta, S., Datta, S.,& Pihur,V. (2021). Package cIVAlid. *Repository Cran*.
- Dewan Perwakilan Rakyat Republik Indonesia (DPR RI). “Referensi Indikator Kesejahteraan”. Publikasi DPR RI diakses dari <http://berkas.dpr.go.id/> ,tanggal 28 Oktober 2021.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. 5th ed. London: John Wiley & Sons.
- Govender, P., & Sivakumar, V. (2020). Application of K-Means and Hierarchical Clustering Techniques for Analysis of Air Pollution: A review (1980–2019). *Atmospheric Pollution Research*, 11(1), 40–56. <https://doi.org/10.1016/j.apr.2019.09.009>.
- Hair, J., Black, W., Babin, B., & Anderson, R. (2013). *Multivariate Data Analysis*. 7th ed . Edinburgh : Pearson.
- Hennig, J., Meila, M., Murtagh, F., & Rocci, R. (2016). *Handbook of Cluster Analysis*. USA: CRC Press.
- Izzati,R.A. “Situasi Kemiskinan Selama Pandemi”. Publikasi Research Institute diakses dari <http://www.smereu.or.id/> , tanggal 04 Januari 2022.
- Johnson, R., & Wichern, D. (2014). *Applied Multivariat Statistical Analysis*. 6th ed. United States of America : Pearson Education.
- Kassambara, A. (2017). *Practical Guide To Cluster Analysis in R*. 1st ed. sthd.
- King, R. S. (2015). *Cluster Analysis and Data Mining An Introduction*. Virginia: