ANALISIS *CLUSTER* DALAM PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN VARIABEL PENYAKIT MENULAR MENGGUNAKAN METODE *COMPLETE LINKAGE*, *AVERAGE LINKAGE* DAN *WARD*

SKRIPSI

UIN SUNAN AMPEL
S U R A B A Y A

Disusun Oleh

RAFIKA VERIANI

H72216064

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA

2020

ANALISIS *CLUSTER* DALAM PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN VARIABEL PENYAKIT MENULAR MENGGUNAKAN METODE *COMPLETE LINKAGE*, *AVERAGE LINKAGE* DAN *WARD*

SKRIPSI

Diajukan guna memenuhi salah satu persyaratan untuk memperoleh
gelar Sarjana Matematika (S.Mat) pada Program Studi Matematika

UIN SUNAN AMPEL
S U R A B A Y A

Disusun oleh
RAFIKA VERIANI
H72216064

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA

2020

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama	:	RAFIKA VERIANI
NIM	:	H72216064
Program Studi	:	Matematika
Angkatan	:	2016

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "ANALISIS *CLUSTER* DALAM PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN VARIABEL PENYAKIT MENULAR MENGGUNAKAN METODE *COMPLETE LINKAGE*, *AVERAGE LINKAGE* DAN *WARD*". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 09 Maret 2020

Yang menyatakan,

METERAI TEMPEL
TGL 20
A74A9AHF494091016
6000
ENAM RIBU RUPIAH

RAFIKA VERIANI
NIM. H72216064

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama	:	RAFIKA VERIANI
NIM	:	H72216064
Judul Skripsi	:	ANALISIS *CLUSTER* DALAM PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN VARIABEL PENYAKIT MENULAR MENGGUNAKAN METODE *COMPLETE LINKAGE*, *AVERAGE LINKAGE* DAN *WARD*

telah diperiksa dan disetujui untuk diujikan.

Surabaya, 09 Maret 2020

Pembimbing

Nurissaidah Ulinnuha, M.Kom

NIP. 199011022014032004

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama	:	RAFIKA VERIANI
NIM	:	H72216064
Judul Skripsi	:	ANALISIS *CLUSTER* DALAM PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN VARIABEL PENYAKIT MENULAR MENGGUNAKAN METODE *COMPLETE LINKAGE*, *AVERAGE LINKAGE* DAN *WARD*

Telah dipertahankan di depan Tim Penguji
pada tanggal 11 Maret 2020

Mengesahkan,
Tim Penguji

Penguji I

Nurissaidah Ulinnuha, M.Kom
NIP. 199011022014032004

Penguji II

Aris Fanani, M.Kom
NIP. 198701272014031002

Penguji III

Yuniar Farida, MT
NIP. 197905272014032002

Penguji IV

Putroue Keumala Intan, M.Si
NIP. 198805282018012001

Mengetahui,
Plt. Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya

Dr. Hj. Evi Fatimatur Rusydiyah, M.Ag
NIP. 197312272005012003

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN
Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Rafika Veriani
NIM : H72216064
Fakultas/Jurusan : Sains dan Teknologi / Matematika
E-mail address : rafikaveriani.rv@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :
☑ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (……)
yang berjudul :

ANALISIS *CLUSTER* DALAM PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN VARIABEL PENYAKIT MENULAR MENGGUNAKAN METODE *COMPLETE LINKAGE, AVERAGE LINKAGE* DAN *WARD*

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara ***fulltext*** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 11 Maret 2020

Penulis

(Rafika Veriani)

ABSTRAK

ANALISIS *CLUSTER* DALAM PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN VARIABEL PENYAKIT MENULAR MENGGUNAKAN METODE *COMPLETE LINKAGE*, *AVERAGE LINKAGE* DAN *WARD*

Penyakit adalah salah satu indikator dalam indeks pembangunan manusia bidang kesehatan. Mengingat bahwa pembangunan bidang kesehatan di Indonesia sedang mengalami beban ganda dimana penyakit menular masih menjadi masalah yang belum dapat diselesaikan, dan masih terdapat penyakit menular yang awalnya masih mampu dikendalikan kini muncul kembali. Hal ini seharusnya mendapatkan perhatian lebih khususnya bagi Dinas Kesehatan maupun Kementerian Kesehatan Republik Indonesia mengenai penyebaran penyakit menular ataupun penyakit tidak menular. Salah satu upaya yang dapat dilakukan adalah dengan membentuk suatu pengelompokan provinsi dalam suatu kelompok yang memiliki karakteristik yang sama dengan maksud memberikan informasi terkait dengan kesehatan pada masing-masing provinsi. Pada penelitian ini bertujuan untuk membentuk suatu *cluster* provinsi di Indonesia berdasarkan variabel jenis penyakit menggunakan metode *Complete Linkage*, *Average Linkage*, dan *Ward*. Ukuran jarak yang digunakan dalam penelitian ini adalah jarak *Euclidean* dan *Squared Euclidean*, dan untuk menentukan cara kerja metode yang terbaik dengan melihat dari nilai simpangan baku dalam kelompok (S_w) yang minimum, simpangan baku antar kelompok (S_b) yang maksimum, dan rasio S_w terhadap S_b yang minimum. Hasil analisis *cluster* yang terbaik adalah dengan metode *Ward* menggunakan 6 *cluster* dengan nilai S_w sebesar 0,18405, S_b sebesar 2,12284 serta rasio S_w terhadap S_b sebesar 0,08670.

Kata kunci: penyakit, *Complete Linkage*, *Avarage Linkage*, *Ward*, jarak *Euclidean*, *Squared Euclidean*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ABSTRACT

CLUSTER ANALYSIS IN THE GROUPING OF PROVINCES IN INDONESIA BASED ON INFECTIOUS DISEASE VARIABLES USING THE *COMPLETE LINKAGE*, *AVERAGE LINKAGE* AND *WARD* METHODS

Disease is one of the indicators human development index in the health sector. Considering that development in the health sector in Indonesia is suffering double burden of disease. Infectious diseases are still a problem that cannot be resolved, and there are still infectious diseases that were initially still able to be controlled now reappear. This should get more attention especially for the Office of Health and the Ministry of Health of the Republic of Indonesia regarding the spread of infectious or non-communicable diseases. One effort that can be done is to form a grouping of provinces in a group that has the same characteristics, the purpose of providing information related to health in each province. This study aims to form a cluster of provinces in Indonesia based on disease type variables using the *Complete Linkage*, *Average Linkage*, and *Ward* methods. The size of the distance used in this study is the distance of is *Euclidean* and *Squared Euclidean*, and to determine how the best method works by looking at the minimum standard deviation in the clusters (S_w) value, the maximum standard deviation between clusters (S_b) value, and the minimum S_w to S_b ratio. The best cluster analysis results are the Ward method using 6 clusters with a value of S_w of 0,18405, S_b of 2,12284, and a ratio of S_w to S_b of 0,08670.

Keywords: disease, *Complete Linkage*, *Average Linkage*, *Ward* methods, *Euclidean*, *Squared Euclidean* distances

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR ISI

HALAMAN JUDUL . . . . . **i**
HALAMAN PERNYATAAN KEASLIAN . . . . . **ii**
LEMBAR PERSETUJUAN PEMBIMBING . . . . . **iii**
PENGESAHAN TIM PENGUJI SKRIPSI . . . . . **iv**
LEMBAR PERNYATAAN PUBLIKASI . . . . . **v**
ABSTRAK . . . . . **vi**
ABSTRACT . . . . . **vii**
DAFTAR ISI **1**
DAFTAR TABEL **4**
DAFTAR GAMBAR **5**
DAFTAR LAMBANG **6**
I PENDAHULUAN **7**
1.1. Latar Belakang Masalah 7
1.2. Rumusan Masalah 12
1.3. Tujuan Penelitian 12
1.4. Manfaat Penelitian 13
1.5. Batasan Masalah 13
1.6. Sistematika Penulisan 14
II TINJAUAN PUSTAKA **16**
2.1. Definisi Kesehatan 16
2.2. Analisis Multivariat 16
2.3. Analisis *Cluster* 17
2.4. Standarisasi Data 18
2.5. Uji Multikolinearitas 18
2.6. Ukuran Jarak 20
2.6.1. Jarak *Euclidean* 20
2.6.2. Jarak *Squared Euclidean* 21

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.7. Metode *Hierarchical Clustering* 21
2.8. Metode Pautan Lengkap *(Complete Linkage)* 23
2.9. Metode Pautan Rata-Rata *(Average Linkage)* 24
2.10. Metode Ward 24
2.11. Simpangan Baku 29
2.12. Interpretasi *Cluster* 30
III METODE PENELITIAN 32
3.1. Jenis Penelitian 32
3.2. Teknik Pengumpulan Data 32
3.3. Variabel Penelitian 32
3.4. Tahap Penelitian 33
IV HASIL DAN PEMBAHASAN 37
4.1. Deskripsi Data 37
4.2. Standarisasi Data 38
4.3. Uji Multikolinearitas 38
4.4. Proses Analis *Cluster* Metode *Complete Linkage* 42
4.5. Proses Analisis *Cluster* Menggunakan Metode *Average Linkage* . . 45
4.6. Proses Analisis *Cluster* Menggunakan Metode *Ward* 48
4.7. Pemilihan Metode Terbaik 56
4.7.1. Simpangan Baku pada Metode *Complete Linkage* dan *Average Linkage* 57
4.7.2. Simpangan Baku pada Metode *Ward* 59
4.7.3. Simpangan Baku Keseluruhan *Cluser* 61
V PENUTUP 66
5.1. Simpulan 66
5.2. Saran 67
DAFTAR PUSTAKA 68
A Data Jumlah Kasus Penyakit di Indonesia 71
B Data Standarisasi 72
C Jarak *Euclidean* Antar Dua Objek 73

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

D Jarak *Squared Euclidean* Antar Dua Objek 81
E SSE Antar Dua Objek 89
F Jumlah *Cluster* dan Anggota Metode *Complete Linkage* 97
G Jumlah *Cluster* dan Anggota Metode *Average Linkage* 98
H Jumlah *Cluster* dan Anggota Metode *Ward* 99
I Rata-Rata 3 *Cluster* Metode *Complete Linkage* dan *Average Linkage* . 100
J Rata-Rata 3 *Cluster* Metode *Ward* 101
K Skrip Program R 102

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR TABEL

4.1 Perhitungan Uji Multikolinearitas Variabel Tuberkulosis dan HIV . 39
4.2 Perhitungan Uji Multikolinearitas Semua Variabel 41
4.3 Perbandingan Rasio Simpangan Baku *Complete Linkage* dan *Average Linkage* 61
4.4 Perbandingan Rasio Simpangan Baku *Ward* 61
4.5 Jumlah *Cluster* dan Anggotanya Metode Ward 62

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR GAMBAR

2.1 Contoh Dendogram 22

3.1 Diagram Alir Metode 33

4.1 Jumlah Kasus Penyakit di Indonesia Tahun 2018 37

4.2 Dendogram *Complete Linkage* 45

4.3 Dendogram *Average Linkage* 48

4.4 Dendogram *Ward* 55

4.5 Ilustrasi Peta Hasil *Cluster* 63

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR LAMBANG

z_i : standarisasi variabel data

x_i : data ke-i

$\bar{x}$: rata-rata seluruh data tiap variabel

s : standar deviasi

R^2 : koefisien determinasi

d_{ij} : jarak *Euclidean* antara obyek ke-i dan obyek ke-j

r_{ij} : jarak *Squared Euclidean* antara obyek ke-i dan obyek ke-j

p : jumlah variabel *cluster*

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

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

d_{ab} : jarak antara objek a dalam *cluster* (IJ) dan objek b dalam *cluster* K

N_{IJ} : jumlah objek pada *cluster* (IJ)

N_K : jumlah objek pada *cluster* K

S_w : simpangan baku dalam *cluster*

S_k : simpangan baku *cluster* ke-k

K : banyak *cluster* yang terbentuk

$\bar{X}$: rataan seluruh *cluster*

$\bar{X}_k$: rataan *cluster* ke-k

S_b : simpangan baku antar *cluster*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Kesehatan adalah salah satu indikator dalam indeks pembangunan manusia. Mengingat bahwa pembangunan bidang kesehatan di Indonesia sedang dihadapkan pada beban ganda dimana penyakit menular masih menjadi masalah yang belum dapat diselesaikan, dan masih terdapat penyakit menular yang awalnya masih mampu dikendalikan kini muncul kembali dengan penyebaran tanpa mengenal batas daerah maupun batas negara. Di sisi lainnya kasus penyakit tidak menular juga semakin meningkat, akibat gaya hidup serta penyakit-penyakit degeneratif. Pemicu terjadinya kecenderungan ini adalah berubahnya gaya hidup, perubahan tersebut mempengaruhi penentuan makanan yang condong lebih menyukai makanan yang siap saji. Padahal makanan tersebut banyak mengandung lemak dan garam yang tinggi, namun kandungan seratnya rendah (Azizah , 2011). Bukan hanya kecenderungan gaya hidup, hal-hal lain yang dapat menghambat kehidupan khususnya di bidang kesehatan diantaranya adalah jumlah penduduk yang semakin meningkat hal ini dapat menyebabkan keterbatasan akan lahan khususnya pada lahan pemukiman, berkurangnya ketersediaan air bersih, menurunnya kualitas air karena limbah, dan meningkatnya pencemaran udara akibat dari pemanfaatan transportasi kota.

Kesehatan merupakan salah satu hal yang dicari oleh semua orang. Tidak ada orang di dunia ini yang menginginkan hidup tidak sehat, namun terkadang

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

datangnya bibit penyakit tidak disangka-sangka. Orang yang terlihat sehat tanpa diduga memiliki penyakit pula. Kesehatan juga merupakan kebutuhan pokok manusia yang sifatnya mutlak. Bagi manusia kesehatan seharusnya sangat penting untuk diperhatikan. Di saat seseorang mengidap atau terserang suatu penyakit maka akan menghambat aktivitas yang positif. Terkadang seseorang akan merasa sangat menjaga kesehatannya saat terserang penyakit. Saat tubuh tidak menunjukkan tanda-tanda adanya keletihan dan penyakit, biasanya seseorang tersebut mengabaikan kesehatannya dengan tidak memberikan waktu untuk istirahat yang cukup bagi tubuhnya. Berdasarkan konteks tersebut sebenarnya, Rasulullah SAW telah mengingatkan kepada umat Islam untuk menjaga kesehatan melalui sabdanya yang diriwayatkan oleh Imam Bukhari:

نِعْمَتَانِ مَغْبُونٌ فِيهِمَا كَثِيرٌ مِنَ النَّاسِ ، الصِّحَّةُ وَالْفَرَاغُ

Artinya: ” *Ada dua nikmat yang sering dilupakan oleh kebanyakan manusia, yaitu nikmat waktu sehat dan waktu senggang*”

Pada hadist diatas menjelaskan bahwa dalam Islam menganjurkan umatnya supaya lebih memperhatikan kesehatan karena kesehatan diri adalah prasyarat untuk meraih kebahagiaan hidup di dunia maupun di akhirat kelak. Mengingat bahwa kesehatan memiliki peran yang sangat penting dalam kehidupan manusia maka sebab itu perlu dilakukan upaya menjaga kesehatan tubuh sejak dini yang berkesinambungan sejak usia bayi masih di dalam kandungan, balita, dewasa, hingga tua. Oleh sebab itu dalam menjaga kesehatan tubuh perlu memperhatikan kondisi lingkungan, gaya hidup sehat, serta gizi yang baik.

Sepuluh tahun terakhir ini Indonesia juga menghadapi masalah *triple burden diseases*. Maksudnya Indonesia masih dilanda penyakit infeksi, penyakit

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

tidak menular (PTM) dan penyakit yang seharusnya sudah teratasi selain itu penyakit menular juga masih menjadi masalah. Hal tersebut ditandai dengan sering terjadi KLB (kejadian luar biasa) beberapa penyakit menular tertentu, munculnya kembali beberapa penyakit menular lama (*re-emerging diseases*), serta munculnya penyakit-penyakit menular baru. Diantara penyakit menular yang sering menelan korban jiwa yaitu salah satunya HIV-AIDS. Di Asia Tenggara Indonesia menduduki posisi tertinggi untuk laju penularan HIV-AIDS. Indonesia juga menempati posisi ketiga sebagai negara dengan penularan kasus HIV-AIDS di wilayah Asia Pasifik setelah India dan China (Ilo & Who , 2018).

Beberapa masalah kesehatan mengenai penyakit menular maupun tidak menular yang ada di Indonesia ini menunjukkan bahwa masyarakat Indonesia belum menyadari tentang pentingnya kesehatan. Hal ini seharusnya mendapatkan perhatian lebih, khususnya bagi Dinas Kesehatan maupun Kementerian Kesehatan Republik Indonesia mengenai penyebaran penyakit menular ataupun penyakit tidak menular. Oleh karena itu, diharapkan semua pihak harus lebih menjaga kesehatan diri. Salah satu upaya yang mampu dilakukan yakni dengan membentuk suatu pengelompokan provinsi ke dalam suatu kelompok yang memiliki kemiripan atau karakteristik yang hampir sama. Hal ini bermaksud untuk memberikan informasi terkait dengan kesehatan pada masing-masing provinsi. Salah satu cara yang dapat dilakukan dalam pengelompokan ini yaitu menggunakan analisis *cluster*.

Pada penelitian sebelumnya telah dilakukan penelitian perbandingan kinerja metode *Complete Linkage* dan *Average Linkage* dalam menentukan hasil analisis *cluster* pada produksi palawija Provinsi Kalimantan Timur oleh Silvia Ningsih, Sri Wahyuningsih, & Yuki N.N (2016). Penelitian tersebut menunjukkan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

hasil bahwa nilai rasio simpangan baku pada metode *Average Linkage* yaitu 0,056 lebih kecil daripada nilai rasio simpangan baku pada metode *Complete Linkage* yaitu 0,083. Berdasarkan hasil tersebut menunjukkan bahwa metode *Average linkage* lebih baik dari metode *Complete Linkage* (Ningsih, Wahyuningsih, & Nasution , 2016). Kemudian pada penelitian lain yang dilakukan oleh Febrina Laraswati (2014) yang membandingkan kinerja metode *Complete Linkage*, *Average Linkage*, *K-Means* dalam menentukan hasil analisis *cluster*. Jarak yang digunakan dalam penelitian tersebut adalah jarak *Euclidean* pada metode *Complete Linkage* dan *K-Means*, serta jarak *Squared Euclidean* pada metode *Average Linkage*. Berdasarkan penelitian tersebut diperoleh hasil bahwa metode *Complete Linkage* dan *Average Linkage* lebih baik dibandingkan dengan metode *K-Means* (Febrina , 2014).

Selain itu penelitian lainnya juga dilakukan oleh Sofya Laeli (2014) dengan judul analisis *cluster* dengan *Average Linkage Method* dan *Wards Method* untuk data responden nasabah asuransi jiwa unit link. Jarak yang digunakan pada kedua metode tersebut adalah *Squared Euclidean*. Berdasarkan penelitian tersebut diperoleh bahwa metode *Average Linkage* memiliki kinerja yang lebih baik daripada metode Ward. Dimana metode *Average Linkage* mempuyai nilai rasio simpangan baku yang lebih kecil yaitu 0,486, dibandingkan dengan nilai rasio simpangan baku metode Ward yaitu 0,710 (Laeli , 2014).

Penelitian yang lain dilakukan oleh Lisda, Ika Purnamasari, & Fidia D.T.A (2018) dengan judul penerapan metode *Complete Linkage* dan Metode *Hierarchical Clustering Multiscale Bootstrap* (Studi Kasus: Kemiskinan Di Kalimantan Timur Tahun 2016). Berdasarkan penelitian tersebut menunjukkan hasil bahwa metode *Complete Linkage* memiliki kinerja yang baik dibandingkan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

metode *Hierarchical Clustering Multiscale Bootstrap* (Ramadhani , 2018). Penelitian yang lainnya dilakukan oleh Satria Fiqih (2016) dengan judul Perbandingan Kinerja Metode *Ward* dan *K-Means* dalam Menentukan Cluster Data Mahasiswa Permohonan Beasiswa. Pada penelitian tersebut diperoleh hasil bahwa metode Ward memiliki kinerja yang lebih baik dibandingkan dengan metode *K-Means*. Berdasarkan penelitian tersebut diperoleh nilai rasio simpangan baku (Sw/Sb) pada metode Ward yaitu 0.5346668% lebih kecil dari metode *K-Means* yang memiliki rasio simpangan baku (S_w/S_b) sebesar 0,831525302 % (Satria & Aziz , 2016).

Berdasarkan pada uraian tersebut, penulis tertarik dalam melakukan penelitian mengenai analisis *cluster* dengan mengambil kasus penyakit yang sering menyerang kesehatan masyarakat di Indonesia diantaranya adalah Tuberkulosis, HIV, AIDS, Kusta, Campak, Difteri, DBD (Demam berdarah Dengue), Pneumonia pada balita, dan Malaria. Judul dalam penelitian ini adalah " Analisis *Cluster* dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode *Complete Linkage*, *Average Linkage*, dan *Ward*". Ketiga metode yang digunakan tersebut bertujuan untuk mengelompokkan kemiripan suatu objek dalam *cluster* yang sama, oleh sebab itu dibutuhkan sebuah ukuran dalam mengetahui seberapa mirip atau tidaknya obyek-obyek tersebut. Dalam mengukur kemiripan atau kesamaan antar obyek dapat dihitung menggunakan ukuran asosiasi, ukuran korelasi, jarak *Cityblock*, jarak *Chebychev*, dan jarak *Euclidean*. Pada penelitian ini penulis menggunakan jarak *Euclidean* pada metode *Complete Linkage*, *Average Linkage*, dan *Squared Euclidean* pada metode *Ward* dalam mengukur kesamaan antar obyek. Untuk pemilihan metode pengelompokan yang memiliki kinerja terbaik antara metode *Complete Linkage*,

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Avarage Linkage dan *Ward* dapat dilihat dari nilai simpangan baku dalam kelompok (S_w) yang minimum dan nilai simpangan baku antar kelompok (S_b) yang maksimum.

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Metode apakah yang terbaik antara metode *Complete Linkage*, *Average Linkage* dan *Ward* dilihat dari nilai simpangan baku dalam kelompok dan simpangan baku antar kelompok pada data jumlah kasus penyakit menular di Indonesia pada tahun 2018?
2. Bagaimana hasil cluster yang terbentuk dengan menggunakan metode terbaik antara *Complete Linkage*, *Average Linkage* dan *Ward* pada data jumlah kasus penyakit menular di Indonesia pada tahun 2018?

1.3. Tujuan Penelitian

Tujuan pada penelitian ini adalah sebagai berikut:

1. Untuk membandingkan metode *Complete Linkage*, *Average Linkage* dan *Ward* apabila dilihat dari simpangan baku dalam kelompok serta nilai simpangan baku antar kelompok dengan simpangan baku tersebut akan diperoleh metode yang terbaik.
2. Untuk membentuk sebuah cluster menggunakan metode terbaik dengan data jumlah kasus penyakit menular di Indonesia pada tahun 2018.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1.4. Manfaat Penelitian

1. Bagi Penulis :

   Hasil penelitian ini dapat digunakan untuk memperdalam wawasan tentang analisis *cluster* dan dalam menganalisis tingkat kesehatan berdasarkan variabel penyakit menular dengan beberapa metode pada analisis *cluster* .

2. Bagi Pembaca :

   Penelitian ini dapat membantu untuk menambah ilmu pengetahuan mengenai analisis *cluster* serta metode yang terbaik dalam pengelompokan tingkat kesehatan berdasarkan variabel penyakit menular di Indonesia.

3. Bagi Kementerian Kesehatan Indonesia :

   Penelitian ini dapat membantu untuk mengetahui tingkat kesehatan masyarakat di setiap provinsi yang ada di Indonesia agar dapat melakukan pencegahan dan pengendalian penyakit menular di tiap Provinsi.

1.5. Batasan Masalah

1. Data yang digunakan adalah jumlah kasus penyakit yang sering menyerang masyarakat di Indonesia pada tahun 2018 seperti Tuberkulosis, HIV, AIDS, Kusta, Campak, Difteri, DBD (Demam berdarah Dengue), Pneumonia pada balita, dan Malaria.

2. Menggunakan jarak *Euclidean* dan *Squared Euclidean* dalam perhitungan kemiripan antar obyek.

3. Menggunakan metode *Complete Linkage*, metode *Average Linkage*, dan *Ward* dalam melakukan analisis *cluster*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4. Menggunakan simpangan baku dalam *cluster* (S_w) serta simpangan baku antar *cluster* (S_b) dalam menentukan metode yang memiliki kinerja yang terbaik
5. Jumlah cluster yang diuji coba adalah dua sampai enam *cluster*.

1.6. Sistematika Penulisan

Sistematika penulisan yang digunakan pada penyusunan skripsi ini adalah sebagai berikut:

1. BAB I: PENDAHULUAN

   Pada bab ini memaparkan tentang latar belakang penulisan skripsi, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

2. Bab II: TINJAUAN PUSTAKA

   Pada bab ini membahas mengenai konsep yang dijadikan dasar teori seperti kesehatan, analisis multivariat, analisis *cluster*, proses analisis *cluster* pemilihan metode dalam analisis cluster menggunakan metode *Complete Linkage*, *Avarage Linkage*, dan Ward, pemilihan metode terbaik menggunakan simpangan baku, dan interpretasi *cluster*.

3. Bab III: METODE PENELITIAN

   Pada bab ini memaparkan tentang jenis penelitian, teknik pengumpulan data, variabel penelitian, dan teknik analisis data.

4. Bab IV: HASIL DAN PEMBAHASAN

   Pada bab ini memaparkan tentang deskripsi hasil penelitian mengenai analisis *cluster* Provinsi di Indonesia berdasarkan variabel jenis penyakit menular menggunakan metode *Complete Linkage*, *Avarage Linkage*, dan *Ward*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

5. Bab IV: PENUTUP

   Pada bab ini memaparkan tentang deskripsi simpulan yakni ringkasan materi dan saran yang bersifat membangun bagi penelitian selanjutnya.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB II

TINJAUAN PUSTAKA

2.1. Definisi Kesehatan

Kesehatan merupakan keadaan yang sejahtera dari badan, jiwa serta sosial dimana setiap orang hidup dengan produktif baik secara sosial maupun ekonomis. Setiap orang sangat penting dalam menjaga kesehatannya terutama dalam upaya menanggulangi dan mencegah gangguan kesehatan yang membutuhkan pemeriksaan, perawatan dan pengobatan (Fauzia & Paramita , 2016). Saat seseorang tidak benar-benar menjaga kesehatan seringkali virus penyakit akan datang menghampiri. Diantara beberapa jenis penyakit menular yang sering menyerang kesehatan manusia diantaranya adalah DBD, HIV, Tuberkulosis (TBC), AIDS, Campak, Difteri, Pneumonia, Kusta, dan Malaria.

2.2. Analisis Multivariat

Analisis multivariat berasal dari kata multi dan variate. Berdasarkan dua kata tersebut maka analisis multivariat ini digunakan dalam teknik statistika tertentu untuk menganalisis banyak variabel yakni secara simultan. Variabel di dalam analisis multivariat dibagi menjadi variabel dependen dan variabel independen. Data dari variabel dependen serta independen dalam analisis multivariat terbagi menjadi dua jenis data yakni kuantitatif dan kualitatif. Variabel kuantitatif merupakan variabel yang tersaji dalam bentuk angka. Contoh variabel kuantitatif misalnya berat badan seseorang, keuntungan perusahaan, harga mobil, dan lain sebagainya. Sedangkan variabel kualitatif merupakan variabel yang tersaji

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

bukan dalam bentuk angka. Oleh sebab itu variabel ini sering kali disebut dengan variabel yang mempunyai data nonmetrik. Contoh dari variabel kualitatif seperti tingkat pendidikan, jenis kelamin, dan lain sebagainya (Ramadhani , 2018).

Analisis multivariate ditujukan dalam melakukan analisis terhadap data yang mempunyai beberapa variabel dimana variabel-variabel tersebut berkorelasi satu sama lain. Umumnya analisis multivariat terbagi menjadi dua yaitu metode dependensi dan metode independensi (Laeli , 2014). Ciri dari metode dependensi adalah adanya satu atau beberapa variabel yang memiliki fungsi sebagai variabel tergantung dan variabel bebas, Metode-metode yang tergolong kelompok metode dependen yaitu *Multiple Regression Analysis* (Analisis Regresi Berganda) *Multiple Regression Analysis* (Analisis Regresi Berganda), *Multiple Discriminant Analysis* (Analisis Diskriminan Berganda), *Multivariate Analysis of Variance* (Analisis Multivariat Varians), *Canonical Correlation Analysis* (Analisis Korelasi Kanonikal) (Soraya , 2011). Sedangkan ciri dari metode interdependensi yaitu seluruh variabelnya memiliki sifat independen. Berikut ini yang termasuk dalam metode interdependensi adalah analisis *cluster*, analisis faktor, dan *multidimensional scaling* (Laeli , 2014).

2.3. Analisis *Cluster*

Analisis *cluster* adalah teknik analisis statistik yang digunakan untuk menempatkan kumpulan obyek ke dalam dua grup bahkan lebih berdasarkan pada karakteristik kesamaan-kesamaan obyek. Analisis cluster merupakan teknik analisis yang bertujuan untuk memilih obyek dalam beberapa kelompok dimana antara kelompok satu dengan kelompok lainnya memiliki sifat yang berbeda. Dalam analisis cluster tiap kelompok bersifat homogen antar anggota dalam kelompok atau variasi obyek dalam satu kelompok yang terbentuk sekecil

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

mungkin. Secara keseluruhan solusi *cluster* tergantung pada variabel yang digunakan sebagai dasar dalam menilai suatu kesamaan. Pengurangan dan penambahan variabel yang terkait dapat mempengaruhi hasil analisis *cluster*. Umumnya analisis *cluster* dapat dikatakan sebagai suatu proses dalam menganalisa baik atau tidaknya proses pembentuk *cluster*.

2.4. Standarisasi Data

Standarisasi data dapat dilakukan jika pada variabel yang diobservasi terdapat perbedaan ukuran satuan yang cukup besar. Perbedaan satuan yang cukup mencolok akan menyebabkan perhitungan pada analisis *cluster* tidak benar-benar valid. Oleh sebab itu hendaknya dilakukan proses standarisasi data melalui transformasi pada data aktual sebelum melakukan analisis yang lebih lanjut. Transformasi dalam bentuk z skor adalah seperti di bawah ini :

$$z = \frac{x_i - \bar{x}}{s} \tag{2.1}$$

Keterangan :

x_i = data ke-i

$\bar{x}$ = rata-rata data

s = standar deviasi

2.5. Uji Multikolinearitas

Terdapat asumsi-asumsi yang harus diperhatikan saat melakukan analisis *cluster*, salah satunya adalah uji Multikolinieritas. Istilah Multikolinieritas mula-mula ditemukan oleh Ragnar Frisch. Multikolinearitas adalah suatu masalah dalam teknik multivariat karena akan menyebabkan dampak yang tidak baik.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Multikolinieritas merupakan adanya suatu hubungan linear yang pasti atau sempurna diantara beberapa atau semua variabel. Hocking (1996) mengemukakan bahwa ada tiga kriteria yang bisa dipakai untuk mengetahui terjadinya multikolinearitas, dan yang paling umum digunakan adalah *Varians Inflation Factor*(*VIF*). Apabila nilai *VIF* lebih besar dari 10 menunjukkan terjadinya multikolinieritas antara variabel-variabel prediktor. Rumus dalam menghitung *Varians Inflation Factor*(*VIF*) sebagai berikut:

$$VIF = \frac{1}{1 - R^2} \tag{2.2}$$

dimana R^2 adalah koefisien determinasi yang dirumuskan dengan:

$$R^2 = \left[\frac{n\Sigma_{i=1}^{n} X_1 X_2 - \Sigma_{i=1}^{n} X_1 \Sigma_{i=1}^{n} X_2}{\sqrt{[n\Sigma_{i=1}^{n}(X_1)^2 - (\Sigma_{i=1}^{v} X_1)^2][n\Sigma_{i=1}^{n}(X_2)^2 - (\Sigma_{i=1}^{v} X_2)^2}} \right]^2 \tag{2.3}$$

Dalam mengetahui adanya multikolinieritas dapat juga dicari menggunakan nilai tolerance dengan rumus :

$$\textit{Tolerance} = \frac{1}{VIF} \tag{2.4}$$

Solusi dalam mengatasi adanya multikolinieritas yaitu dengan mengeluarkan variabel yang tidak berhubungan atau saling terkait. Ada atau tidaknya multikolinieritas antar variabel sangat diperhatikan dalam melakukan analisis *cluster*. Hal ini karena dapat berpengaruh, sehingga variabel-variabel yang bersifat multikolinieritas perlu dipertimbangkan dengan teliti. Jika asumsi multikolinieritas terpenuhi maka jarak *Euclidean* dapat digunakan dalam menentukan ukuran kemiripan (Sholiha , 2015).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.6. Ukuran Jarak

Tujuan dari analisis *cluster* adalah mengelompokkan obyek kemiripan suatu obyek dalam *cluster* yang sama. Oleh karena dibutuhkan beberapa ukuran guna mengetahui seberapa kemiripan serta perbedaan obyek-obyek tersebut. Dalam mengukur kesamaan antar obyek ada tiga metode yang bisa digunakan yaitu ukuran asosiasi, ukuran korelasi, serta ukuran jarak. Pada penelitian kali ini peneliti menggunakan ukuran jarak(Febrina , 2014).

Metode ukuran jarak digunakan pada data yang memiliki skala metrik. Ukuran jarak ini sebetulnya adalah ukuran ketidakmiripan, dimana untuk jarak yang besar memiliki kesamaan yang sedikit sedangkan untuk jarak yang kecil menjelaskan bahwasannya suatu obyek akan lebih mirip dengan obyek yang lainnya. Untuk penelitian kali ini ukuran jarak yang dipakai yaitu menggunakan ukuran jarak *Euclidean*, dan *Squared Euclidean* (Febrina , 2014).

2.6.1. Jarak *Euclidean*

Jarak *Euclidean* adalah besar suatu jarak pada garis lurus yang menghubungkan antar objek. Sebagai contoh ukuran ketidaksamaan atau jarak antar obyek ke-i dan obyek ke-j dapat disimbolkan dengan d_{ij} dan untuk variabel ke-k dengan k=1,...,p. nilai d_{ij} didapatkan dari perhitungan pada jarak Euclidean dengan rumus :

$$d_{ij} = \sqrt{(\Sigma_{k=1}^{p} x_{ik} - x_{jk})^2} \tag{2.5}$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

keterangan :

d_{ij} = jarak *Euclidean* dari obyek ke-i dan obyek ke-j

p = jumlah variabel cluster

x_{ik} = nilai objek ke-i pada variabel ke-k

x_{jk} = nilai objek ke-j pada variabel ke-k

2.6.2. Jarak *Squared Euclidean*

Jarak *Squared Euclidean* adalah variasi dari jarak *Euclidean*. Perbedaan dari kedua jarak yaitu jika jarak *Euclidean* diakarkan, sedangkan jarak *Squared Euclidean* tidak terdapat akar. Dengan rumus sebagai berikut :

$$r_{ij} = (\Sigma_{k=1}^{p} x_{ik} - x_{jk})^2 \tag{2.6}$$

keterangan :

r_{ij} = jarak *Squared Euclidean* dari obyek ke-i dan obyek ke-j

p = jumlah variabel cluster

x_{ik} = nilai objek ke-i pada variabel ke-k

x_{jk} = nilai objek ke-j pada variabel ke-k

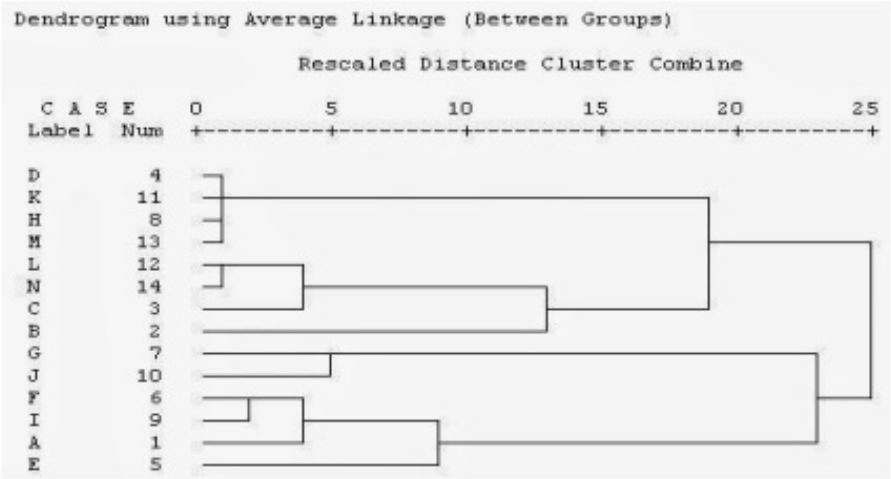
2.7. Metode *Hierarchical Clustering*

Metode pengelompokan data dapat dilakukan menggunakan dua metode yakni metode hierarki serta metode non hierarki. Namun dalam penelitian kali ini penulis hanya membahas mengenai metode hierarki. Metode hierarki merupakan Aglomerasi (penggabungan) atau Divisive (pemecahan). Pada metode aglomerasi setiap penelitian pada awalnya diasumsikan sebagai *cluster* tersendiri yang mana terdapat *cluster* sesuai jumlah penelitiannya. Selanjutnya untuk dua *cluster* yang

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

memiliki kesamaan terdekat digabungkan menjadi *cluster* baru, dengan begitu jumlah *cluster* akan berkurang satu untuk tiap tahapan. Sedangkan untuk pemecahan akan dimulai dari satu *cluster* besar yang mengandung seluruh observasi. Selanjutnya untuk observasi yang tidak sama akan dipisahkan dan dibentuk kelompok-kelompok yang lebih kecil. Proses ini akan terus dilakukan sampai setiap observasi terbentuk kelompok sendiri-sendiri (Laeli , 2014).

Metode hierarki pengelompokannya dimulai dengan dua objek atau lebih yang memiliki kesamaan terdekat. Selanjutnya perhitungan akan diteruskan ke objek lainnya yang memiliki kedekatan selanjutnya atau kedua. Begitupun seterusnya hingga *cluster* akan membentuk Dendogram atau semacam pohon dengan terdapat hierarki (tingkatan) yang jelas antara objeknya, mulai yang paling mirip hingga yang paling tidak mirip (Hartini , 2014). Dendogram digunakan dalam menunjukkan anggota dari *cluster* yang ada apabila ingin menentukan jumlah *cluster* yang seharusnya akan dibentuk. Contoh dari Dendogram tertera pada gambar 2.1



Gambar 2.1 Contoh Dendogram

Jika akan membaca dendogram maka cara membacanya dari sisi sebelah

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

kiri ke kanan. Dengan melihat dendogram pada gambar diatas, apabila ingin membentuk dua *cluster*, dapat dilihat bahwa untuk cluster pertama mempunyai anggota dari obyek D hingga obyek B, serta pada cluster yang kedua mempunyai anggota dari obyek G hingga obyek E (Febrina , 2014). Terdapat lima metode Aglomerasi atau gabungan untuk pembentukan cluster yakni metode pautan tunggal *(Single Linkage)*, pautan lengkap *(Complete Linkage)*, pautan rata-rata *(Average Linkage)*, metode Ward (Wards method), dan metode pusat *(Centroid)*.

2.8. Metode Pautan Lengkap *(Complete Linkage)*

Completed linkage method (metode pautan lengkap) dikembangkan oleh Horn (1943) dan Sorensen (1948). Metode ini sering disebut juga metode maksimum (Latan , 2014). Metode *Complete Linkage* berdasar pada jarak maksimum. Pengukuran jarak antara *cluster* satu dengan yang lainnya berdasarkan pada obyek yang memiliki jarak paling jauh. Semua obyek dalam suatu *cluster* saling dikaitkan satu sama lain pada suatu kesamaan minimum ataupun dengan jarak maksimum. Dalam metode *Complete Linkage*, ditentukan menggunakan jarak terjauh antara dua objek pada gerombol yang berbeda (*Furthest Neighbor*). Dalam perhitungan awal, nilai minimum dalam $D = d_{ij}$ dicari terlebih dulu serta digabungkan pada obyek-obyek yang bersesuaian, contohnya I dan K, guna memperoleh *cluster* (IJ). Jarak antara *cluster* lain K dan (IJ), dihitung menggunakan cara :

$$d_{(IJ)K} = max\ \{d_{IK}, d_{JK}\} \tag{2.7}$$

d_{IK} dan d_{JK} adalah jarak terjauh diantara anggota *cluster-cluster* I dan K serta *cluster-cluster* J dan K (Johnson et,al. , 1996).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.9. Metode Pautan Rata-Rata *(Average Linkage)*

Metode *Average Linkage* merupakan proses *clustering* (*Average Distance*) atau pada jarak rata-rata antar obyeknya. Untuk metode *Average Linkage*, jarak diantara dua *cluster* dapat diasumsikan sebagai jarak rata-rata dari semua anggota dalam satu *cluster* dengan semua anggota dalam *cluster* lainnya. Metode ini memiliki tujuan untuk meminimumkan rataan jarak semua pasangan observasi dan dua gerombol yang di gabungkan. Metode ini dianggap lebih stabil diantara metode *hierarchical* yang lainnya. Tahapan pertama yang harus dilakukan pada metode ini sama seperti metode-metode sebelumnya yakni mencari jarak minimum (Ningsih, Wahyuningsih, & Nasution , 2016).

$$d_{(IJ)K} = \frac{\Sigma_a \Sigma_b d_{ab}}{N_{(IJ)} N_K} \tag{2.8}$$

keterangan :

d_{ab} = jarak antara obyek a dalam cluster (IJ) dan obyek b dalam cluster K

$N_{(IJ)}$ = jumlah obyek pada cluster (IJ)

N_K = jumlah obyek pada cluster K

2.10. Metode Ward

Metode Ward merupakan metode varians yang umum digunakan. Metode varians bertujuan untuk mendapatkan *cluster* yang memiliki varians internal cluster yang sekecil mungkin. Pada metode Ward rata-rata untuk setiap *cluster* dihitung (Ramadhani , 2018). Kemudian, dihitung jarak *squared euclidean* antara setiap obyek dan nilai rata-ratanya, lalu jarak tersebut dihitung semua, setiap tahap, dua *cluster* yang mempunyai kenaikan *sum of squares* dalam *cluster* yang terkecil

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

digabungkan (Simarora , 2005).

Metode Ward merupakan suatu metode pembentukan *cluster* yang didasari oleh hilangnya informasi akibat penggabungan obyek menjadi *cluster*. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat pada mean *cluster* untuk setiap pengamatan. *Error sum of squares* (SSE) digunakan sebagai fungsi obyektif. Dua obyek akan digabungkan jika mempunyai fungsi obyektif terkecil diantara kemungkinan yang ada. SSE hanya dapat dihitung jika *cluster* memiliki elemen lebih dari satu objek. SSE *cluster* yang hanya memiliki satu objek adalah nol. Metode Ward ini dihitung berdasarkan persamaan berikut rumus dalam menghitung SSE (Rencher , 2002):

$$SSE = \Sigma_{i=1}^{N}(X_i - \bar{X})'(X_i - \bar{X}) \tag{2.9}$$

Keterangan :

X_i = vektor kolom yang entrinya nilai objek ke-i dengan i yaitu 1,2,3,...N

$\bar{X}$ = vektor kolom yang entrinya rata-rata objek dalam *cluster*

N = banyaknya objek

Apabila U adalah *cluster* I, maka jumlah kuadrat galat di dalam Cluster adalah:

$$SSE_U = \Sigma_{i=1}^{n_U}(X_i - \bar{X}_U)'(X_i - \bar{X}_U) \tag{2.10}$$

Jika V adalah *cluster* II, maka jumlah kuadrat galat di dalam Cluster adalah:

$$SSE_V = \Sigma_{i=1}^{n_V}(X_i - \bar{X}_V)'(X_i - \bar{X}_V) \tag{2.11}$$

Jika UV merupakan *cluster* yang diperoleh dengan menggabungkan *cluster* U dan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

V, maka jumlah kuadrat didalam *cluster* adalah sebagai berikut:

$$SSE_{UV} = \Sigma_{i=1}^{n_{UV}} (X_i - \bar{X}_{UV})'(X_i - \bar{X}_{UV}) \tag{2.12}$$

Dimana :

SSE_U, SSE_V, SSE_{UV} = Jumlah kuadrat galat Cluster U,V dan UV

$\bar{X}_U, \bar{X}_V, \bar{X}_{UV}$ = Vektor kolom yang entrinya rata-rata dari Cluster U,V

n_U, n_V, n_{UV} = Banyaknya *Cluster* pada objek U,V dan UV, Sehingga:

$$\bar{X}_U = \Sigma_{i-1}^{n_U} \frac{X_i}{n_U}$$

$$\bar{X}_V = \Sigma_{i-1}^{n_V} \frac{X_i}{n_V}$$

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

Dalam menggabungkan dua *cluster* U dan V dengan cara meminimalkan peningkatan SSE yang didefinisikan sebagai jarak antara Cluster U dan Cluster V yaitu seperti di bawah ini :

$$I_{UV} = SSE_{UV} - (SSE_U + SSE_V) \tag{2.13}$$

Dari persamaan diatas dapat ditunjukkan bahwa peningkatan I_{UV}

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

mempunyai bentuk ekuivalen seperti berikut ini:

$$
\begin{aligned}
I_{UV} &= SSE_{UV} - (SSE_U + SSE_V) \\
I_{UV} &= n_U(\bar{X}_U - \bar{X}_{UV})'(\bar{X}_U - \bar{X}_{UV}) + n_V(\bar{X}_V - \bar{X}_{UV})'(\bar{X}_V - \bar{X}_{UV}) \\
&= n_U\left(\bar{X_U} - \left(\frac{n_U\bar{X_U}n_V\bar{X_V}}{n_U+n_V}\right)\right)'\left(\bar{X_U} - \left(\frac{n_U\bar{X_U}n_V\bar{X_V}}{n_U+n_V}\right)\right) + \\
&n_V\left(\bar{X_V} - \left(\frac{n_U\bar{X_U}n_V\bar{X_V}}{n_U+n_V}\right)\right)'\left(\bar{X_V} - \left(\frac{n_U\bar{X_U}n_V\bar{X_V}}{n_U+n_V}\right)\right) \\
&= n_U\left(\frac{(n_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+n_V)\bar{X_U} - n_U\bar{X_U} - n_V\bar{X_V}}{n_U+n_V}\right) + \\
&n_V\left(\frac{(n_U+n_V)\bar{X_V} - n_U\bar{X_U} - n_V\bar{X_V}}{n_U+n_V}\right)' \\
&\left(\frac{(n_U+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_V} - \bar{X_U})}{n_U+n_V}\right)'\left(\frac{n_U(\bar{X_V} - \bar{X_U})}{n_U+n_V}\right) \\
&= \frac{n_U n_V{}' n_U n_V}{(n_U+n_V)'(n_U+n_V)}\left((\bar{X_U} - \bar{X_V})'(\bar{X_U} - \bar{X_V})\right) + \\
&\frac{n_V n_U{}' n_V n_U}{(n_U+n_V)'(n_U+n_V)}\left((\bar{X_V} - \bar{X_U})'(\bar{X_V} - \bar{X_U})\right) \\
&= \frac{n_U n_V{}' n_U n_V + n_V n_U{}' n_V n_U}{(n_U+n_V)'(n_U+n_V)}\left((\bar{X_U} - \bar{X_V})'(\bar{X_U} - \bar{X_V})\right) \\
&= \frac{n_U(n_V n_V') + n_V(n_U n_U')}{(n_U+n_V)'(n_U+n_V)}\left((\bar{X_U} - \bar{X_V})'(\bar{X_U} - \bar{X_V})\right) \\
&= \frac{n_U n_V(n_V{}') + n_V n_U(n_U{}')}{n_U n_U{}' + n_U n_V{}' + n_V n_U{}' + n_V n_V{}'}\left((\bar{X_U} - \bar{X_V})'(\bar{X_U} - \bar{X_V})\right) \\
&= \frac{n_U n_V(n_V{}' + n_U{}')}{n_U(n_U{}' + n_V{}') + n_V(n_U{}' + n_V{}')}\left((\bar{X_U} - \bar{X_V})'(\bar{X_U} - \bar{X_V})\right) \\
&= \frac{n_U n_V(n_V{}' + n_U{}')}{n_U + n_V(n_V{}' + n_U{}')}\left((\bar{X_U} - \bar{X_V})'(\bar{X_U} - \bar{X_V})\right)
\end{aligned}
$$

$$
= \frac{n_U n_V}{n_U + n_V}\left((\bar{X_U} - \bar{X_V})'(\bar{X_U} - \bar{X_V})\right) \tag{2.14}
$$

Dari Persamaan (2.14) tersebut untuk meminimalkan peningkatan SSE adalah ekuivalen dengan melakukan peminimalan jarak antar objek. Apabila U terdiri dari X_i serta V terdiri dari X_j, maka SSE_U dan SSE_V yaitu bernilai nol. Dalam

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

menentukan jarak dengan metode Ward menggunakan persamaan jarak antara objek dengan rumus seperti di bawah ini :

$$\begin{aligned} I_{ij} = SSE_{ij} &= \frac{1}{2}(X_i - X_j)(X_i - X_j)' \\ &= \frac{1}{2}d^2(X_i, X_j) \\ &= \frac{1}{2}\Sigma_{k=1}^{p}(X_{ik} - X_{jk})^2 \end{aligned} \tag{2.15}$$

Dimana :

I_{ij} = jarak antara objek i dan j

X_i = nilai rata-rata objek i

X_j = nilai rata-rata objek j

$d^2(X_i, X_j)$ = jarak kuadrat *Euclidean* antara objek I dan J

k = 1, 2, 3,... p

p = jumlah variabel *cluster*

Berdasarkan Persamaan (2.14) jarak antara objek UV dan objek W dengan menggunakan metode Ward yaitu seperti berikut ini:

$$I_{(UV)W} = \frac{n_U + n_W}{n_{UV} + n_W} I_{UW} + \frac{n_V n_W}{n_{UV} + n_W} I_{VW} - \frac{n_W}{n_{UV} + n_W} I_{UV} \tag{2.16}$$

Dimana :

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

I_{UW} = Jarak antara *cluster*U dan W

I_{VW} = Jarak antara *cluster* V dan W

n_U, n_V, n_W = Banyak objek pada *cluster* U, V, W

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.11. Simpangan Baku

Dalam mengetahui metode mana yang memiliki cara kerja yang paling baik, bisa menggunakan rata-rata simpangan baku dalam *cluster* (S_w) serta simpangan baku antar cluster (S_b) (Fadlina, , 2015). Untuk rumus S_w adalah sebagai berikut:

$$S_w = \frac{1}{K}\Sigma_{k=1}^{K} S_k \tag{2.17}$$

Keterangan:

K = banyaknya cluster yang terbentuk

S_k = simpangan baku cluster ke-k.

Apabila diberikan cluster c_k dimana k=1,...,p, dan setiap *cluster* memiliki anggota x_i, dimana i=1,...,n dan n adalah jumlah anggota dari setiap cluster , dan $\bar{x_k}$ adalah rata-rata dari *cluster* k maka untuk mencari nilai simpangan baku ke-k (s_k) digunakan rumus berikut (Barakbah, A.,R. , 2004):

$$S_k = \sqrt{\frac{1}{n-1}\Sigma_{k=1}^{n}(x_i - \bar{x_k})^2} \tag{2.18}$$

Apabila terdapat rata-rata variabel dalam setiap cluster k $(\bar{x_k})$ maka komponen dari setiap cluster akan berbeda, dan simpangan baku antar kelompok (s_b) dapat dirumuskan sebagai berikut :

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$S_b = \left[\frac{1}{K-1}\Sigma_{k=1}^{K}(\bar{X}_k - \bar{X})^2\right]^{\frac{1}{2}} \tag{2.19}$$

Keterangan:

K = Banyaknya kelompok yang terbentuk

$\bar{X}_k$ = Rataan cluster ke-k

$\bar{X}$ = Rataan seluruh cluster

Metode pengelompokan yang baik memiliki nilai simpangan baku antar kelompok (S_b) yang maksimum dan nilai simpangan baku dalam kelompok (S_w) yang minimum. Nilai rasio S_w terhadap S_b yang semakin kecil dapat dikatakan metode tersebut mempunyai kinerja yang baik karena memiliki tingkat homogenitas yang tinggi di dalam *cluster* (Satria & Aziz , 2016). Rasio S_w terhadap S_b dapat dihitung dengan rumus di bawah ini:

$$S = \frac{S_w}{S_b} \tag{2.20}$$

2.12. Interpretasi *Cluster*

Dalam mengetahui bagaimana profil dari setiap kelompok dengan melakukan rata-rata untuk tiap variabelnya. Oleh karena itu penting untuk menginterpretasi *cluster*. Misal dilakukan pembentukan 3 *cluster*, maka dalam melakukan interpretasi profil *cluster*, pernyataan yang digunakan adalah seperti di bawah ini :

1. *Cluster* yang memiliki rata-rata paling rendah akan digolongkan sebagai kelompok provinsi yang mempunyai kesehatan baik.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. *Cluster* yang memiliki rata-rata lebih tinggi dibandingkan cluster rata-rata paling rendah maka akan digolongkan sebagai kelompok provinsi yang mempunyai kesehatan kurang baik atau cukup rawan penyakit .

3. *Cluster* yang memiliki rata-rata paling tinggi akan digolongkan sebagai kelompok provinsi yang rawan penyakit

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB III

METODE PENELITIAN

3.1. Jenis Penelitian

Penelitian ini merupakan penelitian yang bersifat kualitatif, karena data yang digunakan pada penelitian ini adalah jumlah penderita penyakit Tuberkulosis, HIV, AIDS, Kusta, Campak, Difteri, DBD, Pneomonia pada balita, dan Malaria di Indonesia. Dilihat dari tujuannya penelitian ini termasuk penelitian komperatif karena membandingkan tiga teori yaitu *Complete Linkage*, *Avarage Linkage* dan *Ward*.

3.2. Teknik Pengumpulan Data

Penelitian ini menggunakan observasi sebagai teknik pengumpulan datanya. Observasi merupakan teknik pengumpulan data dengan cara mencatat segala macam informasi untuk kepentingan penelitian dan dibuktikan selama penelitian.

3.3. Variabel Penelitian

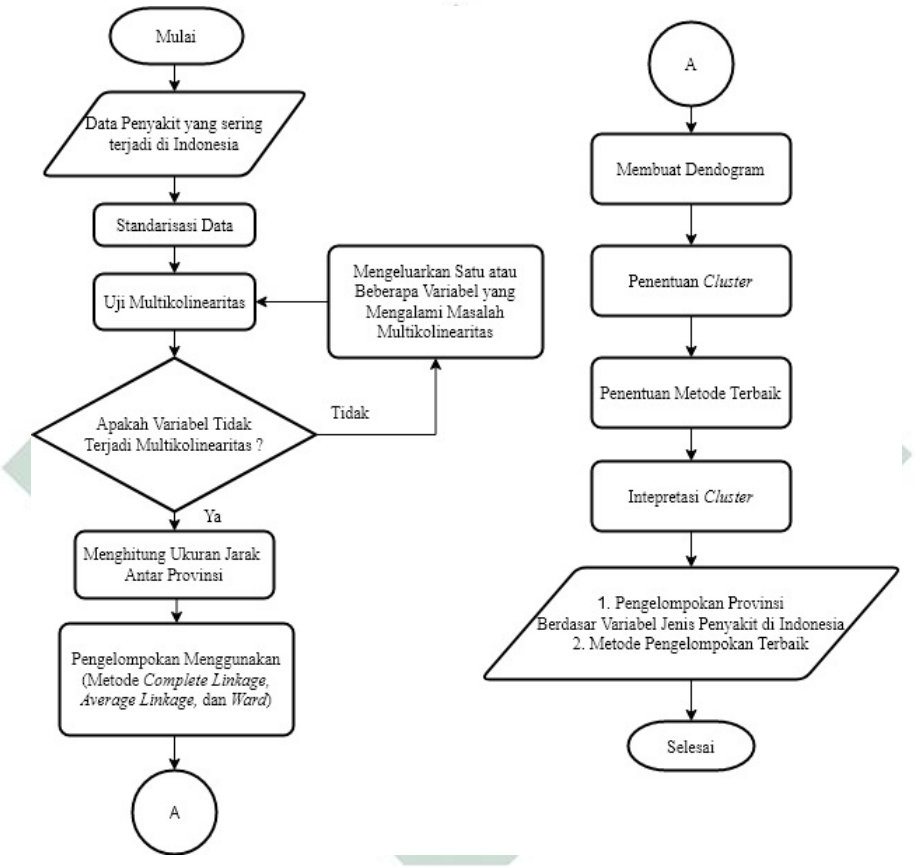
Dalam penelitian ini, variabel penelitiannya adalah jumlah dari beberapa kasus penyakit yang menyerang masyarakat yaitu penyakit Tuberkulosis (X1), HIV (X2), AIDS (X3), Kusta (X4), Campak (X5), Difteri (X6), DBD (Demam berdarah Dengue) (X7), Pneomonia pada balita (X8), dan Malaria (X9) di Indonesia pada tahun 2018. Data tersebut diperoleh dari Kementerian Kesehatan Republik Indonesia yang dapat diakses melalui Publikasi Profil Kementerian

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Kesehatan, 2018.

3.4. Tahap Penelitian

Tahapan penelitian ini digambarkan melalui *flowchart* yang ditunjukkan oleh Gambar 3.1.



Gambar 3.1 Diagram Alir Metode

Berdasarkan Gambar 3.1 adapun langkah-langkah yang harus dilakukan dalam penelitian ini adalah sebagai berikut :

1. Masukkan data penyakit Tuberkulosis, HIV, AIDS, Kusta, Campak, Difteri, DBD (Demam berdarah Dengue), Pneomonia pada balita, dan Malaria.
2. Melakukan standarisasi data menggunakan *z-score* karena data yang digunakan di dalam penelitian ini memiliki satuan yang bervariasi atau berbeda yang

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ditunjukkan oleh persamaan (2.1) .

3. Melakukan uji multikolinieritas guna mengetahui suatu data terjadi korelasi antar variabel atau tidak. Uji multikolinieritas dilakukan dengan mencari nilai Variance Inflation Factor(VIF) yang ditujukkan pada persamaan (2.2). Jika nilai VIF < 10 atau nilai *tolerance* > 0.10, maka variabel tersebut memiliki pengaruh terhadap multikolinieritas dengan variabel bebas lainnya.

4. Melakukan pengelompokan setiap provinsi menggunakan metode *Complete Linkage*. Langkah-langkah pengelompokan menggunakan metode *Complete Linkage* adalah seperti berikut:

   a. Mengukur jarak antar objek (dua provinsi) dalam penelitian dengan menggunakan rumus jarak *Euclidean* yang ditunjukkan pada persamaan (2.5)

   b. Menentukan jarak yang terdekat antar objek (provinsi)

   c. Melakukan perhitungan ulang jarak dengan metode *Complete Linkage* yang ditunjukkan pada persamaan (2.7)

   d. Menentukan jumlah anggota *cluster* serta anggotanya dengan membentuk dendogram sesuai dengan *cluster*.

   e. Melakukan pembentukan *cluster* menggunakan *metode Complete Linkage*

   f. Menghitung simpangan baku dalam kelompok pada persamaan (2.17), simpangan baku antar kelompok pada persamaan (2.19) dan rasio simpangan baku dalam kelompok terhadap simpangan baku antar kelompok dengan persamaan (2.20)

5. Melakukan pengelompokan menggunakan metode *Average Linkage* dengan langkah-langkah seperti berikut:

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

a. Mengukur jarak antar objek (dua provinsi) dalam penelitian dengan menggunakan rumus jarak *Euclidean* yang ditunjukkan pada persamaan (2.5)

b. Menentukan jarak yang terdekat antar objek (provinsi)

c. Melakukan perhitungan ulang jarak dengan metode *Average Linkage* yang ditunjukkan pada persamaan (2.8)

d. Menentukan jumlah anggota *cluster* serta anggotanya dengan memberikan dendogram sesuai dengan *cluster* yang terbentuk

e. Melakukan pembentukan *cluster* menggunakan *metode Average Linkage*

f. Menghitung simpangan baku dalam kelompok pada persamaan (2.17) simpangan baku antar kelompok pada persamaan (2.19) dan rasio simpangan baku dalam kelompok terhadap simpangan baku antar kelompok dengan persamaan (2.20)

6. Melakukan pengelompokan menggunakan metode *Ward* dengan langkah-langkah seperti berikut:

a. Mengukur jarak antar objek (dua provinsi) dalam penelitian dengan menggunakan rumus jarak *Squared Euclidean* yang ditunjukkan pada persamaan (2.6)

b. *Cluster* pertama dibentuk dengan cara menentukan dua objek yang mempunyai nilai SSE terkecil menggunakan persamaan (2.15)

c. Menentukan jarak yang terdekat antar objek (provinsi).

d. Melakukan perhitungan ulang jarak dengan metode *Ward* yang ditunjukkan pada persamaan (2.16)

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

e. Mengulangi langkah (b)-(d) sampai semua objek bergabung menjadi satu *cluster*.

f. Menentukan jumlah *Cluster* dan anggotanya dengan Metode Ward

g. Melakukan pembentukan *Cluster* dengan Metode Ward.

h. Menghitung simpangan baku dalam kelompok pada persamaan (2.17), simpangan baku antar kelompok pada persamaan (2.19) dan rasio simpangan baku dalam kelompok terhadap simpangan baku antar kelompok dengan persamaan (2.20)

7. Menentukan metode terbaik dengan melihat hasil dari simpangan baku

8. Melakukan intepretasi *Cluster*

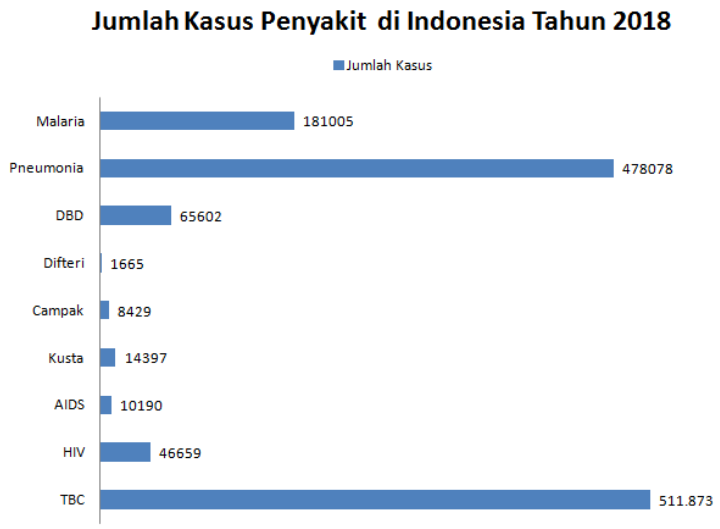
digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB IV

HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Sebelum melakukan proses perhitungan data, terlebih dahulu yang harus dilakukan adalah mengumpulkan data. Data yang digunakan dalam penelitian ini secara tidak langsung diperoleh dari publikasi Kementerian Kesehatan Republik Indonesia, yaitu mengenai jumlah kasus penyakit yang sering terjadi di Indonesia tahun 2018. Adapun data terkait 34 provinsi dengan 9 jenis kasus penyakit tertera pada Lampiran A. Data 34 provinsi di Indonesia digunakan sebagai objek pengamatan, dengan karakteristiknya adalah jumlah kasus penyakit yang sering terjadi. Kasus penyakit yang dimaksud adalah Tuberkulosis, HIV, AIDS, Kusta, Campak, Difteri, DBD, Pneomonia pada balita, dan Malaria. Jumlah kasus penyakit di Indonesia tahun 2018 secara menyeluruh disajikan pada Gambar 4.1.



Gambar 4.1 Jumlah Kasus Penyakit di Indonesia Tahun 2018

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Berdasarkan Gambar 4.1 terlihat bahwa jenis kasus penyakit di Indonesia yang paling banyak terjadi adalah Tuberkulosis dengan jumlah kasus sebesar 511.783 kasus sedangkan kasus penyakit yang terendah adalah Difteri sebesar 1.665 kasus.

4.2. Standarisasi Data

Data yang digunakan dalam penelitian ini yang tertera pada Lampiran A, mempunyai jangkauan yang terpaut jauh antara variabel. Oleh sebab itu harus dilakukan standarisasi data. Proses perhitungan standarisasi data yang dimaksud dengan melakukan transformasi pada data asli menggunakan *z-score* berdasarkan Persamaan 2.1 sebagai berikut :

Tuberkulosis = Variabel X_1

$$\bar{x} = \frac{8.145 + 32651 + ... + 10813}{34} = 15.055$$

$$s = \sqrt{\frac{\Sigma_{i=1}^{34}(x_i - \bar{x})^2}{34 - 1}}$$

$$s = \sqrt{15.105.985.499}$$

$$s = 21.395,26$$

$$z_{Aceh} = \frac{x - \bar{x}}{s} = \frac{8.145 - 15.055}{21.395,26} = -0,32297$$

begitupun seterusnya.

Adapun hasil standarisasi seluruh data tersebut dapat dilihat dalam Lampiran B.

4.3. Uji Multikolinearitas

Untuk mengetahui terjadi atau tidaknya multikolinearitas, dapat dilihat dari nilai VIF dan *tolerance* pada masing-masing variabel independen. Suatu data dikatakan terbebas dari gejala multikolinearitas apabila nilai VIF < 10 dan nilai *tolerance* > 0.10. Nilai VIF dapat dihitung menggunakan Persamaan (2.2) dan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

sedangkan *tolerance* pada Persamaan (2.4). Contoh hasil perhitungan uji multikolinearitas variabel Tuberkulosis dan HIV disajikan dalam Tabel 4.1.

Tabel 4.1 Perhitungan Uji Multikolinearitas Variabel Tuberkulosis dan HIV

Provinsi	X_1	X_2	$X_1 * X_2$	$(X_1)^2$	$(X_2)^2$
Aceh	-0,3229	-0,5811	0,1877	0,1043	0,3377
Sumatra Utara	0,8224	0,2991	0,24602	0,6764	0,0895
Sumatra Barat	-0,2010	-0,3572	0,0718	0,0404	0,1276
Riau	-0,1832	-0,3978	0,0729	0,0336	0,1582
Jambi	-0,5028	-0,5376	0,2704	0,2528	0,2891
Sumatera Selatan	0,1429	-0,4125	-0,0589	0,0204	0,1702
Bengkulu	-0,5502	-0,6040	0,3323	0,3027	0,3648
Lampung	0,0241	-0,4049	-0,0097	0,0006	0,1639
Kepulauan Bangka Belitung	-0,5937	-0,5572	0,3308	0,3525	0,3105
Kepulauan Riau	-0,4376	-0,1619	0,0709	0,1915	0,0263
DKI Djakarta	0,9902	2,6367	2,6109	0,9805	6,9522
Jawa Barat	3,9420	1,8200	7,1746	15,5404	3,3123
Jawa Tengah	2,4308	1,9226	4,6735	5,9088	3,6964
DI Yogyakarta	-0,5272	-0,2574	0,1357	0,2779	0,0663
Jawa Timur	1,9345	3,4539	6,6818	3,7424	11,9296
Banten	0,6423	-0,0183	-0,0117	0,4125	0,0003
Bali	-0,5114	0,4003	-0,2047	0,2615	0,1603
NTB	-0,4028	-0,5634	0,2269	0,1622	0,3174
NTT	-0,3884	-0,2532	0,0983	0,1508	0,0641
Kalimantan Barat	-0,3966	-0,3248	0,1288	0,1573	0,1055
Kalimantan Tengah	-0,5378	-0,5968	0,3209	0,2893	0,3562

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Provinsi	X_1	X_2	$X_1 * X_2$	$(X_1)^2$	$(X_2)^2$
Kalimantan Selatan	-0,3200	-0,5133	0,1642	0,1024	0,2635
Kalimantan Timur	-0,4051	-0,1176	0,0476	0,1641	0,0138
Kalimantan Utara	-0,6313	-0,5758	0,3635	0,3985	0,3316
Sulawesi Utara	-0,3865	-0,3902	0,1508	0,1494	0,1522
Sulawesi Tengah	-0,3811	-0,5157	0,1965	0,1453	0,2659
Sulawesi Selatan	0,3913	-0,0947	-0,0371	0,1531	0,0089
Sulawesi Tenggara	-0,5124	-0,6045	0,3097	0,2625	0,3654
Gorontalo	-0,5391	-0,6198	0,3341	0,2906	0,3841
Sulawesi Barat	-0,6057	-0,6427	0,3892	0,3668	0,4130
Maluku	-0,5328	-0,4345	0,2315	0,2838	0,1889
Maluku Utara	-0,6157	-0,5596	0,3445	0,3790	0,3132
Papua Barat	-0,6373	-0,4737	0,3018	0,4061	0,2244
Papua	-0,1983	1,0376	-0,2057	0,0393	1,0766
Jumlah	0	0	25,9399	33	32,999

Terlebih dahulu yang perlu dilakukan dalam mencari nilai VIF dan *tolerance* adalah menghitung koefisien determinasi menggunakan Persamaan (2.3) sebagai berikut:

$$R^2 = \left[\frac{n\Sigma_{i=1}^{n} X_1 X_2 - \Sigma_{i=1}^{n} X_1 \Sigma_{i=1}^{n} X_2}{\sqrt{[n\Sigma_{i=1}^{n}(X_1)^2 - (\Sigma_{i=1}^{v} X_1)^2][n\Sigma_{i=1}^{n}(X_2)^2 - (\Sigma_{i=1}^{v} X_2)^2}}\right]^2$$

$$R^2 = \left[\frac{(34)(25{,}9399) - (0)(0)}{\sqrt{[(34)(3{,}30*(10)) - (0)^2][(34)(32{,}999) - (0)^2]}}\right]^2$$

$$R^2 = (0,78606)^2$$

$$R^2 = 0,61789$$

Kemudian menghitung nilai VIF :

$$VIF = \frac{1}{1 - R^2}$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$VIF = \frac{1}{1 - 0,61789}$$
$$VIF = 2,61704$$

Selanjutnya menghitung nilai *tolerance* :

$$Tolerance = \frac{1}{VIF}$$
$$Tolerance = \frac{1}{2,61704}$$
$$Tolerance = 0,38211$$

Hasil dari perhitungan uji multikolinearitas seluruh variabel yang digunakan dapat dilihat pada Tabel 4.2.

Tabel 4.2 Perhitungan Uji Multikolinearitas Semua Variabel

No	Variabel	VIF	*Tolerance*
1	Tuberkulosis	2,6170	0,3821
2	HIV	2,5974	0,3850
3	AIDS	1,6387	0,6103
4	Kusta	1,0177	0,9826
5	Campak	1,0979	0,9108
6	Difteri	1,0874	0,9196
7	DBD	2,8693	0,3485
8	Pneumonia Balita	1,0114	0,9888
9	Malaria	1,0025	0,9975

Berdasarkan pada Tabel 4.2 dapat dilihat bahwa pada perhitungan uji multikolinearitas untuk semua variabel yang digunakan tidak terdapat data yang mengalami multikolinearitas karena nilai VIF < 10 dan nilai *Tolerance* $> 0,10$.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.4. Proses Analis *Cluster* Metode *Complete Linkage*

Pada metode *Complete Linkage* proses pengelompokannya adalah dengan menggabungkan dua obyek atau lebih yang memiliki jarak paling jauh. Untuk mengukur kemiripan antar dua objek pada metode *Complete Linkage* digunakan jarak *Euclidean*. Data yang digunakan adalah data yang telah dilakukan standarisasi pada Lampiran B. Di bawah ini merupakan contoh perhitungan jarak secara manual antara objek 1 (Aceh) dan 2 (Sumatra Utara), objek 1 (Aceh) dan 3 (Sumatra Barat), objek 2 (Sumatra Utara) dan 3 (Sumatra Barat), objek 2 (Sumatra Utara) dan 4 (Riau), serta objek 3 (Sumatra Barat) dan 4 (Riau). Keempat objek tersebut dibandingkan dengan menggunakan sembilan variabel, dengan tujuan untuk memperoleh dua provinsi yang paling mirip diantara empat provinsi tersebut.

Di bawah ini adalah kemiripan antara provinsi Aceh dan Sumatra Utara (objek 1 dan 2):

$$d_{(1,2)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2 + ... + (x_9 - y_9)^2}$$

$$d_{(1,2)} = \sqrt{(-0,32297 - 0,82242)^2 + ... + (-0,22074 - (-0,14146))^2}$$

$$d_{(1,2)} = \sqrt{32,37316}$$

$$d_{(1,2)} = 5,68974$$

Perhitungan kemiripan antara objek 1 dan 2 menghasilkan jarak *Euclidean* sebesar 5,68974. Kemudian untuk perhitungan kemiripan antara objek 1 dan 3 adalah :

$$d_{(1,3)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2}$$

$$d_{(1,3)} = \sqrt{(-0,32297 - (-0,20103))^2 + ... + (-0,22074 - (-0,21249)^2}$$

$$d_{(1,3)} = \sqrt{24,65689}$$

$$d_{(1,3)} = 4,96557$$

Dari perhitungan tersebut diperoleh jarak *Euclidean* sebesar 4,96557. Selanjutnya

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

perhitungan kemiripan objek 2 dan 3 :

$d_{(2,3)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2}$

$d_{(2,3)} = \sqrt{(0,82242 - (-0,20103))^2 + ... + (-0,14146 - (-0,21249))^2}$

$d_{(2,3)} = \sqrt{4,54045}$

$d_{(2,3)} = 2,13083$

Dari perhitungan tersebut diperoleh jarak *Euclidean* sebesar 2,13083. Kemudian menghitung kemiripan objek 2 dan 4 :

$d_{(2,4)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2}$

$d_{(2,4)} = \sqrt{(0,82242 - (-0,18322))^2 + ... + (-0,14146 - (-0,22303))^2}$

$d_{(2,4)} = \sqrt{6,79286}$

$d_{(2,4)} = 2,60631$

Diperoleh jarak *Euclidean* sebesar 2,60631. Kemudian menghitung kemiripan objek 3 dan 4 :

$d_{(3,4)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_{11} - y_{11})^2}$

$d_{(3,4)} = \sqrt{-0,20103 - (-0,18322))^2 + ... + (-0,21249 - (-0,22303))^2}$

$d_{(3,4)} = \sqrt{0,56826438}$

$d_{(3,4)} = 0,75383$

Dari hasil perhitungan jarak *Euclidean* antara Provinsi Aceh dan Sumatra Utara diperoleh jarak sebesar 4,977473, untuk jarak Provinsi Aceh dan Sumatra Barat sebesar 2,13083, untuk Provinsi Sumatra Utara dan Sumatra Barat sebesar 2,60631, serta untuk Provinsi Sumatra Barat dan Riau sebesar 0,75383. Hal ini menunjukkan bahwa jarak *Euclidean* antara Provinsi Sumatra Barat dan Riau memiliki karakteristik yang lebih mirip dibandingkan dengan Provinsi Aceh dan Sumatra Barat serta Sumatra Utara. Hasil perhitungan jarak *Euclidean* antar provinsi dapat dilihat pada Lampiran C. Untuk membentuk *cluster* baru, dilakukan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dengan menggabungkan jarak antara dua objek yang paling mirip atau mempunyai jarak terkecil. Berdasarkan pada Lampiran C, jarak antar provinsi yang paling kecil sebesar 0,27817 yaitu jarak antara *cluster* objek 28 (Provinsi Sulawesi Tenggara) dengan objek 31 (Provinsi Maluku) yang kemudian kedua provinsi tersebut digabung menjadi satu *cluster*.

Selanjutnya melakukan pembentukan matriks jarak baru menggunakan Persamaan (2.7) berdasarkan Lampiran C. Jarak antar provinsi yang melibatkan pembentukan *cluster* baru adalah objek 28 (Provinsi Sulawesi Tenggara) dan objek 31 (Provinsi Maluku). Dalam menghitung pembentukan matriks jarak baru adalah sebagai berikut ini :

$d_{(28,31)1} = maxd_{(28)1}, d_{(31)1} = maxd_{(5,5326)}, d_{(5,6133)} = 5,6133$

$d_{(28,31)2} = maxd_{(28)2}, d_{(31)2} = maxd_{(2,9424)}, d_{(3,0460)} = 3,0460$

$d_{(28,31)3} = maxd_{(28)3}, d_{(31)3} = maxd_{(1,2721)}, d_{(1,4429)} = 1,4429$

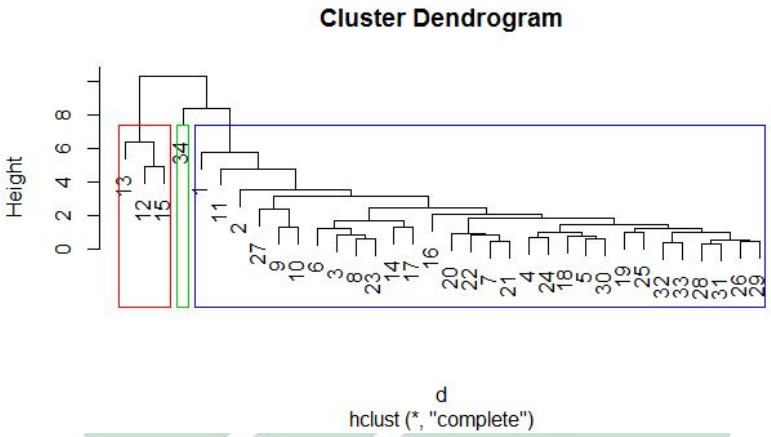
$d_{(28,31)4} = maxd_{(28)4}, d_{(31)4} = maxd_{(0,7374)}, d_{(0,8740)} = 0,8740$

$\vdots$

$d_{(28,31)34} = maxd_{(28)34}, d_{(31)34} = maxd_{(7,1830)}, d_{(7,1510)} = 7,1830$

Begitupun seterusnya perhitungan dilakukan hingga perhitungan matriks jarak dari semua objek yang telah digabungkan pada proses *Agglomeration* dan mendapatkan matriks baru. Proses pengelompokan juga dapat digambarkan dalam bentuk dendogram. Berikut ini adalah dendogram pengelompokan provinsi di Indonesia menggunakan metode *Complete Linkage* jika akan membentuk tiga *cluster* yang disajikan pada Gambar 4.2 :

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 4.2 Dendogram *Complete Linkage*

Saat melakukan analisis *cluster* yang dapat ditunjukkan hanyalah anggota-anggota *cluster* pada banyaknya *cluster* tertentu, namun tidak untuk memutuskan berapa banyak *cluster* yang terbentuk.

4.5. Proses Analisis *Cluster* Menggunakan Metode *Average Linkage*

Untuk mengukur kemiripan antar dua objek pada metode *Average Linkage* digunakan jarak *Euclidean*. Dengan menggunakan data yang telah dilakukan standarisasi pada Lampiran B. Di bawah ini merupakan contoh perhitungan jarak secara manual antara objek 1 (Aceh) dan 2 (Sumatra Utara), objek 1 (Aceh) dan 3 (Sumatra Barat), objek 2 (Sumatra Utara) dan 3 (Sumatra Barat), objek 2 (Sumatra Utara) dan 4 (Riau), serta objek 3 (Sumatra Barat) dan 4 (Riau). Keempat objek tersebut dibandingkan dengan menggunakan sembilan variabel, dengan tujuan untuk memperoleh dua provinsi yang paling mirip diantara empat provinsi tersebut. Di bawah ini adalah kemiripan antara provinsi Aceh dan Sumatra Utara (objek 1 dan 2):

$$d_{(1,2)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2 + ... + (x_9 - y_9)^2}$$

$$d_{(1,2)} = \sqrt{(-0,32297 - 0,82242)^2 + ... + (-0,22074 - (-0,14146))^2}$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$d_{(1,2)} = \sqrt{32,37316}$

$d_{(1,2)} = 5,68974$

Perhitungan kemiripan antara objek 1 dan 2 menghasilkan jarak *Euclidean* sebesar 5,68974. Kemudian untuk perhitungan kemiripan antara objek 1 dan 3 adalah :

$d_{(1,3)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2}$

$d_{(1,3)} = \sqrt{(-0,32297 - (-0,20103))^2 + ... + (-0,22074 - (-0,21249)^2}$

$d_{(1,3)} = \sqrt{24,65689}$

$d_{(1,3)} = 4,96557$

Dari perhitungan tersebut diperoleh jarak *Euclidean* sebesar 4,96557. Selanjutnya perhitungan kemiripan objek 2 dan 3 :

$d_{(2,3)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2}$

$d_{(2,3)} = \sqrt{(0,82242 - (-0,20103))^2 + ... + (-0,14146 - (-0,21249))^2}$

$d_{(2,3)} = \sqrt{4,54045}$

$d_{(2,3)} = 2,13083$

Dari perhitungan tersebut diperoleh jarak *Euclidean* sebesar 2,13083. Kemudian menghitung kemiripan objek 2 dan 4 :

$d_{(2,4)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2}$

$d_{(2,4)} = \sqrt{(0,82242 - (-0,18322))^2 + ... + (-0,14146 - (-0,22303))^2}$

$d_{(2,4)} = \sqrt{6,79286}$

$d_{(2,4)} = 2,60631$

Diperoleh jarak *Euclidean* sebesar 2,60631. Kemudian menghitung kemiripan objek 3 dan 4 :

$d_{(3,4)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_{11} - y_{11})^2}$

$d_{(3,4)} = \sqrt{-0,20103 - (-0,18322))^2 + ... + (-0,21249 - (-0,22303))^2}$

$d_{(3,4)} = \sqrt{0,56826438}$

$d_{(3,4)} = 0,75383$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dari perhitungan jarak *Euclidean* antara Provinsi Aceh dan Sumatra Utara diperoleh jarak sebesar 5,68974, untuk jarak Provinsi Aceh dan Sumatra Barat sebesar 4,96557 untuk Provinsi Sumatra Utara dan Sumatra Barat sebesar 2,13083 untuk Provinsi Sumatra Utara dan Riau sebesar 2,60631 serta untuk Provinsi Sumatra Barat dan Riau sebesar 0,75383. Hal ini menunjukkan bahwa jarak *Euclidean* antara Provinsi Sumatra Barat dan Riau memiliki karakteristik yang lebih mirip dibandingkan dengan Provinsi Aceh dan Sumatra Barat serta Sumatra Utara. Kemudian untuk membentuk *cluster* baru, dilakukan dengan menggabungkan jarak antara dua objek yang paling mirip atau mempunyai jarak terkecil. Berdasarkan pada Lampiran C jarak antar provinsi yang paling kecil sebesar 0,27817 yaitu jarak antar *cluster* objek 28 (Provinsi Sulawesi Tenggara) dan objek 31 (Provinsi Maluku) yang kemudian kedua provinsi tersebut digabung menjadi satu *cluster*.

Selanjutnya melakukan pembentukan matriks jarak baru menggunakan Persamaan (2.8) berdasarkan Lampiran 5. Jarak antar provinsi yang melibatkan pembentukan *cluster* baru adalah objek 28 (Provinsi Sulawesi Tenggara) dan objek 31 (Provinsi Maluku). Dalam menghitung pembentukan matriks jarak baru adalah seperti di bawah ini :

$$d_{(28,31)1} = \frac{d_{(28)1}, d_{(31)1}}{2} = \frac{d_{(5,5326)} + d_{(5,6133)}}{2} = 5,5730$$

$$d_{(28,31)2} = \frac{d_{(28)2}, d_{(31)2}}{2} = \frac{d_{(2,9424)} + d_{(3,0460)}}{2} = 2,9942$$

$$d_{(28,31)3} = \frac{d_{(28)3}, d_{(31)3}}{2} = \frac{d_{(1,7181)} + d_{(2,1210)}}{2} = 1,9191$$

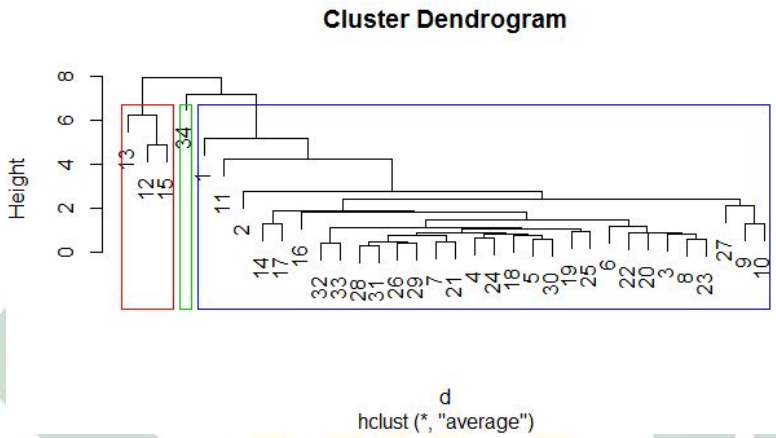
$$d_{(28,31)4} = \frac{d_{(28)4}, d_{(31)4}}{2} = \frac{d_{(0,7374)} + d_{(0,8740)}}{2} = 0,8057$$

$$\vdots$$

$$d_{(28,31)34} = \frac{d_{(28)34}, d_{(31)34}}{2} = \frac{d_{(7,1830)} + d_{(7,1510)}}{2} = 7,1670$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Perhitungan dilakukan begitupun seterusnya hingga perhitungan matriks jarak dari semua objek yang telah digabungkan pada proses *Agglomeration* dan mendapatkan matriks baru. Proses pengelompokan juga dapat digambarkan dalam bentuk dendogram. Berikut ini adalah dendogram pengelompokan provinsi di Indonesia menggunakan metode *Average Linkage* jika akan membentuk tiga *cluster* yang disajikan pada Gambar 4.3 :



Gambar 4.3 Dendogram *Average Linkage*

Saat melakukan analisis *cluster* yang dapat ditunjukkan hanyalah anggota-anggota *cluster* pada banyaknya *cluster* tertentu, namun tidak untuk memutuskan berapa banyak *cluster* yang terbentuk.

4.6. Proses Analisis *Cluster* Menggunakan Metode *Ward*

Untuk mengukur kemiripan antar dua objek pada metode *Ward* digunakan jarak *Squared Euclidean* karena terdapat salah satu proses perhitungan yang membutuhkan hasil dari perhitungan kuadrat *Euclidean* yang tertera pada Persamaan 2.15. Berikut ini merupakan perhitungan jarak secara manual antara provinsi 1 (Aceh) dan 2 (Sumatra Utara), provinsi 1 (Aceh) dan 3 (Sumatra Barat), provinsi 2 (Sumatra Utara) dan 3 (Sumatra Barat), provinsi 2 (Sumatra Utara) dan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4 (Riau), serta provinsi 3 (Sumatra Barat) dan 4 (Riau). Keempat provinsi tersebut dibandingkan dengan menggunakan sembilan variabel, dengan tujuan untuk memperoleh dua provinsi yang paling mirip diantara empat provinsi tersebut. Di bawah ini adalah kemiripan antara provinsi Aceh dan Sumatra Utara (objek 1 dan 2):

$r_{(1,2)} = (x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2$

$r_{(1,2)} = (-0,32297 - 0,82242)^2 + ... + (-0,22074 - (-0,14146))^2$

$r_{(1,2)} = 32,37316$

Untuk perhitungan kemiripan antara objek 1 dan 2 menghasilkan jarak *Squared Euclidean* sebesar 32,37316. Kemudian untuk perhitungan kemiripan antara objek 1 dan 3 adalah :

$r_{(1,3)} = (x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2$

$r_{(1,3)} = (-0,32297 - (-0,20103))^2 + ... + (-0,22074 - (-0,21249)^2$

$r_{(1,3)} = 24,65690$

Dari perhitungan tersebut diperoleh jarak *Squared Euclidean* sebesar 24,65690. Selanjutnya perhitungan kemiripan objek 2 dan 3 :

$r_{(2,3)} = (x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2$

$r_{(2,3)} = (0,82242 - (-0,20103))^2 + ... + (-0,14146 - (-0,21249))^2$

$r_{(2,3)} = 4,54045$

Dari perhitungan tersebut diperoleh jarak *Squared Euclidean* sebesar 4,54045. Kemudian menghitung kemiripan objek 2 dan 4 :

$r_{(2,4)} = (x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2$

$r_{(2,4)} = (0,82242 - (-0,18322))^2 + ... + (-0,14146 - (-0,22303))^2$

$r_{(2,4)}) = 6,79287$

Diperoleh jarak *Squared Euclidean* sebesar 6,79287. Kemudian menghitung kemiripan objek 3 dan 4 :

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$r_{(3,4)} = (x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_9 - y_9)^2$

$r_{(3,4)} = -0,20103 - (-0,18322))^2 + ... + (-0,21249 - (-0,22303))^2$

$r_{(3,4)} = 0,56826$

Dari contoh diatas dapat dilihat bahwa jarak *Squared Euclidean* antara Provinsi Aceh dan Sumatra Utara sebesar 32,37316 , untuk jarak Provinsi Aceh dan Sumatra Barat sebesar 24,65690, untuk Provinsi Sumatra Utara dan Sumatra Barat sebesar 4,54045 serta untuk Provinsi Sumatra Barat dan Riau sebesar 0,56826. Hal ini menunjukkan bahwa jarak *Squared Euclidean* antara Provinsi Sumatra Barat dan Riau memiliki karakteristik yang lebih mirip dibandingkan dengan Provinsi Aceh dan Sumatra Barat serta Sumatra Utara. Hasil perhitungan jarak *Squared Euclidean* antar provinsi dapat dilihat pada Lampiran D. Kemudian dalam membentuk *cluster* baru dengan menggabungkan jarak antara dua objek yang paling mirip atau mempunyai jarak terkecil. Berdasarkan pada Lampiran E jarak antar provinsi yang paling kecil sebesar 0,07738 yaitu jarak antar *cluster* objek 28 (Provinsi Sulawesi Tenggara) dan objek 31 (Provinsi Maluku) yang kemudian kedua provinsi tersebut digabung menjadi satu *cluster*.

Pada proses Agglomerasi dimulai dari dua objek terdekat, dimana jarak antar dua objek tersebut merupakan jarak yang terdekat dari kombinasi 34 objek yang digunakan. Misalnya Provinsi Sulawesi Tenggara dan Maluku dengan menggunakan Persamaan (2.15) sebagai berikut:

$I_{(ij)} \quad = \quad SSE_{(ij)}$

$SSE_{(ij)} = \frac{1}{2}(X_i - X_j)(X_i - X_j)'$

$SSE_{(ij)} = \frac{1}{2}d^2(X_i, X_j)$

$SSE_{(ij)} = \frac{1}{2}\Sigma_{k=1}^{p}(X_{ik} - X_{jk})^2$

$SSE_{(ij)} = \frac{1}{2}(-0,51236 - (-0,53279))^2 + (-0,60448 - (-0,43454))^2 + (-0,39219 -$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$(-0,48672))^2 + (-0,22777 - (-0,19897))^2 + (-0,70264 - (-0,77628))^2 + (-0,52243 - (-0,4997))^2 + (-0,63201 - (-0,78064))^2 + (-0,57204 - (-0,42277))^2 + (-0,19087 - (-0,14273))^2$

$SSE_{(ij)} = 0,03869$

Hasil nilai SSE dua objek lainnya tertera pada Lampiran E. Proses perhitungan diatas hanya menghitung dua kelompok, dimana untuk setiap kelompok terdiri dari satu objek. Selanjutnya yaitu melakukan perhitungan pada jarak suatu *cluster* dengan *cluster* baru yang akan terbentuk. Sebagai contoh adalah objek 28 (Provinsi Sulawesi Tenggara) dan objek 31 (Provinsi Maluku) sebagai berikut:

$I_{(28,31)1} = \frac{(n_{28}+n_1)I_{28,1}+(n_{31}+n_1)I_{31,1}-n_1 I_{28,31}}{n_{28}+n_{31}+n_1}$

$I_{(28,31)1} = \frac{(1+1)30,6099+(1+1)31,5088(1(0,0774)}{1+1+1}$

$I_{(28,31)1} = 41,4383$

Perhitungan Provinsi Sulawesi Tenggara dan Maluku akan dilakukan untuk seluruh provinsi kecuali Provinsi Sulawesi Tenggara dan Maluku itu sendiri. Jarak yang digunakan untuk disatukan menjadi satu *cluster* adalah jarak yang terkecil. Proses tersebut akan terus berlanjut sampai seluruh provinsi bergabung menjadi satu *cluster*.

Pada penelitian kali ini akan mengambil contoh dalam membentuk tiga *cluster*. Dengan *Cluster* I yang beranggotakan Provinsi Aceh (1), DKI Djakarta (11), Jawa Tengah (13), Papua (34). *Cluster* II yang beranggotakan provinsi yaitu Sumatra Utara (2), Sumatera Barat (3), Riau (4), Jambi (5), Sumatera Selatan (6), Bengkulu (7), Lampung (8), Kepulauan Bangka Belitung (9), Kepulauan Riau(10), DI Yogyakarta (14), Banten (16), Bali (17), Nusa Tenggara Barat (18), Nusa Tenggara Timur (19), Kalimantan Barat (20), Kalimantan Tengah (21), Kalimantan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Selatan (22), Kalimantan Timur (23), Kalimantan Utara (24), Sulawesi Utara (25), Sulawesi Tengah (26), Sulawesi Selatan (27), Sulawesi Tenggara (28), Gorontalo (29), Sulawesi Barat (30), Maluku (31), Maluku Utara (32), dan Papua Barat (33) . Untuk *Cluster* III beranggotakan provinsi Jawa Barat (12), serta Jawa Timur (15) maka:

a. SSE Objek *Cluster* I

Nilai untuk SSE objek pada *Cluster* I adalah :

$SSE_I = \Sigma_{i=1}^{n_I}(X_i - \bar{X}_I)'(X_i - \bar{X}_I)$

$SSE_I = (X_1 - \bar{X}_I)^2 + (X_2 - \bar{X}_I)^2 + ... + (X_{36} - \bar{X}_I)^2$

$SSE_I = (-0,32297 - (0,95073))^2 + ... + (5,60547 - (0,95073))^2$

$SSE_I = 87,9631$

b. SSE Objek *Cluster* II

Nilai untuk SSE objek pada *Cluster* II adalah :

$SSE_{II} = \Sigma_{i=1}^{n_{II}}(X_i - \bar{X}_{II})'(X_i - \bar{X}_{II})$

$SSE_{II} = (X_1 - \bar{X}_{II})^2 + (X_2 - \bar{X}_{II})^2 + ... + (X_{252} - \bar{X}_{II})^2$

$SSE_{II} = (0,82242 - (-0,27649))^2 + ... + (0,09034 - (-0,27649))^2$

$SSE_{II} = 32,1829$

c. SSE Objek *Cluster* III

Nilai untuk SSE objek pada *Cluster* III adalah :

$SSE_{III} = \Sigma_{i=1}^{n_{III}}(X_i - \bar{X}_{III})'(X_i - \bar{X}_{III})$

$SSE_{III} = (X_1 - \bar{X}_{III})^2 + (X_2 - \bar{X}_{III})^2 + ... + (X_{18} - \bar{X}_{III})^2$

$SSE_{III} = (3,9421 - 1,9695)^2 + ... + (-0,21084 - 1,9695)^2$

$SSE_{III} = 45,5284$

Dari hasil perhitungan diatas diperoleh bahwa untuk *Cluster* II memiliki nilai

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

terdekat antara dua *Cluster* yang lain, jarak pada *Cluster* II adalah sebesar 32,1829 . Sedangkan untuk jarak paling jauh antara *Cluster* lainnya yaitu *Cluster* I yang mempunyai jarak sebesar 87,9631.

Langkah selanjutnya adalah melakukan pembentukan antar dua *Cluster* yang terbentuk, seperti di bawah ini :

1. Pembentukan Untuk *Cluster* I dan II Nilai untuk SSE objek pada *Cluster* I, II ($SSE_{I,II}$)

   $SSE_{I,II} = \Sigma_{i=1}^{n_{I,II}}(X_i - \bar{X}_{I,II})'(X_i - \bar{X}_{I,II})$

   $SSE_{I,II} = (X_1 - \bar{X}_{I,II})^2 + (X_2 - \bar{X}_{I,II})^2 + ... + (X_{288} - \bar{X}_{I,II})^2$

   $SSE_{I,II} = (-0,32297 - (-0,12309))^2 + ... + (0,09034(-0,12309))^2$

   $SSE_{I,II} = 178,1517$

   Dari perhitungan diatas diperoleh nilai $SSE_{I,II}$ sebesar 178,1517 . Selanjutnya menghitung total jarak antara *Cluster* I dan II sebagai berikut:

   $I_{I,II} = SSE_{I,II} - (SSE_I + SSE_{II})$

   $I_{I,II} = 177,2902 - (87,9631 + 32,1829)$

   $I_{I,II} = 177,2902 - 120,146$

   $I_{I,II} = 57,1442$

   Maka total dari nilai $I_{I,II}$ adalah sebesar 57,1442.

2. Pembentukan Untuk *Cluster* I dan III Nilai untuk SSE objek pada *Cluster* I,III ($SSE_{I,III}$)

   $SSE_{I,III} = \Sigma_{i=1}^{n_{I,III}}(X_i - \bar{X}_{I,III})'(X_i - \bar{X}_{I,III})$

   $SSE_{I,III} = (X_1 - \bar{X}_{I,III})^2 + \cdots + (X_{54} - \bar{X}_{I,III})^2$

   $SSE_{I,III} = (-0,32297 - (1,2903))^2 + \cdots + (-0,21084 - 1,2903)^2$

   $SSE_{I,III} = 144,0039$

   Dari perhitungan diatas diperoleh nilai $SSE_{I,III}$ sebesar 144,0039 . Selanjutnya

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

menghitung total jarak antara *Cluster* I dan III sebagai berikut:

$I_{I,III} = SSE_{I,III} - (SSE_I + SSE_{III})$

$I_{I,III} = 144,0039 - (87,9631 + 45,5284)$

$I_{I,III} = 144,0039 - 133,4915$

$I_{I,III} = 10,5123$

Maka total dari nilai $I_{I,III}$ adalah sebesar 10,5123.

3. Pembentukan Untuk *Cluster* II dan III Nilai untuk SSE objek pada *Cluster* II,III ($SSE_{II,III}$)

   $SSE_{II,III} = \Sigma_{i=1}^{n_{II,III}}(X_i - \bar{X}_{II,III})'(X_i - \bar{X}_{II,III})$

   $SSE_{II,III} = (X_1 - \bar{X}_{II,III})^2 + ... + (X_{270} - \bar{X}_{II,III})^2$

   $SSE_{II,III} = (0,82242 - (-0,1268))^2 + ... + (0,09034 - (-0,1268))^2$

   $SSE_{II,III} = 174,0997$

Dari perhitungan diatas diperoleh nilai $SSE_{II,III}$ sebesar 174,0997. Selanjutnya menghitung total jarak antara *Cluster* II dan III sebagai berikut :

$I_{II,III} = SSE_{II,III} - (SSE_{II} + SSE_{III})$

$I_{II,III} = 174,0997 - (32,1829 + 45,5284)$

$I_{II,III} = 96,3884$

Berdasarkan pada hasil perhitungan diatas dapat diketahui bahwa untuk total jarak dengan nilai jarak paling jauh adalah sebesar 96,3884 yang terjadi pada 30 provinsi, serta untuk total jarak dengan nilai jarak terdekat adalah sebesar 10,5123 yang terjadi pada 6 provinsi. Kemudian jarak antar objek I,II dan III dengan metode Ward adalah seperti di bawah ini :

$I_{(I,III)II} = \frac{n_I+n_{II}}{n_{I,III}+n_{II}} I_{I,II} + \frac{n_{III}+n_{II}}{n_{I,III}+n_{II}} I_{II,III} - \frac{n_{II}}{n_{I,III}+n_{II}} I_{I,III}$

$I_{(I,III)II} = \frac{4+28}{6+28}(57,1442) + \frac{2+30}{6+28}(96,3884) - \frac{6}{6+28}(10,4797)$

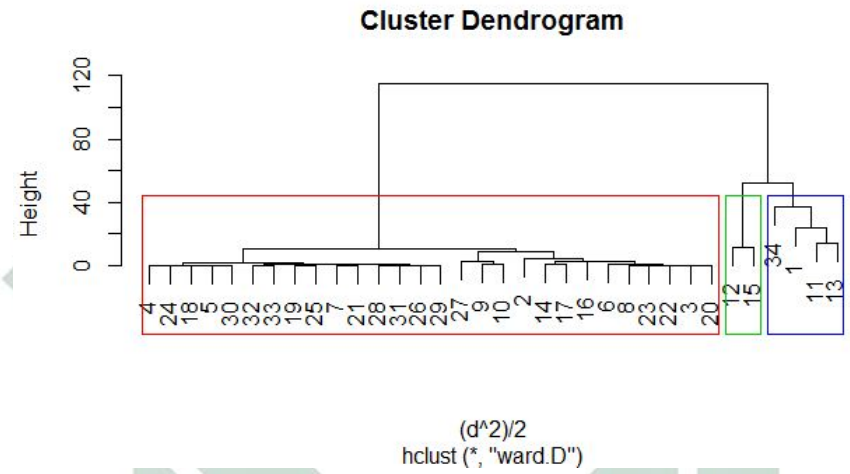
$I_{(I,III)II} = \frac{32}{34}(57,1442) + \frac{32}{34}(96,3884) - \frac{6}{34}(10,4797)$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$I_{(I,III)II} = 53,7828 + 90,7185 - 1,8493$

$I_{(I,III)II} = 142,652$

Jadi jarak yang tergabung dalam *cluster* baru dengan perubahan perhitungan seperti diatas dengan nilai sebesar 142,652. Berikut ini adalah dendogram pengelompokan provinsi di Indonesia menggunakan metode *Ward* yang disajikan pada Gambar 4.4 :



Gambar 4.4 Dendogram *Ward*

Dari Dendogram di atas dapat dijadikan sebagai panduan dalam menentukan objek manakah yang mempunyai kesamaan karakteristik. Dua objek yang mempunyai kesamaan karakteristik akan digambarkan dalam dua titik dengan posisi yang berdekatan. Posisi dari dua objek yang semakin dekat maka menunjukkan dua objek tersebut semakin memiliki kemiripan, dan apabila posisi dari dua titik objek yang semakin jauh maka semakin berbeda karakteristiknya. Berdasarkan analisis hasil dendogram di atas menunjukkan bahwa provinsi yang saling berdekatan adalah :

1. *Cluster* 1 memiliki anggota berjumlah 4 provinsi yaitu Aceh (1), DKI Djakarta

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

(11), Jawa Tengah (13), Papua (34) untuk *cluster* pertama ini mempunyai nilai SSE sebesar 87,9631

2. *Cluster* 2 memiliki anggota berjumlah 28 provinsi yang terdiri dari Sumatra Utara (2), Sumatera Barat (3), Riau (4), Jambi (5), Sumatera Selatan (6), Bengkulu (7), Lampung (8), Kepulauan Bangka Belitung (9), Kepulauan Riau (10), DI Yogyakarta (14), Banten (16), Bali (17), Nusa Tenggara Barat (18), Nusa Tenggara Timur (19), Kalimantan Barat (20), Kalimantan Tengah (21), Kalimantan Selatan (22), Kalimantan Timur (23),Kalimantan Utara (24), Sulawesi Utara (25), Sulawesi Tengah (26), Sulawesi Selatan (27), Sulawesi Tenggara (28), Gorontalo (29), Sulawesi Barat (30), Maluku (31), Maluku Utara (32), Papua Barat (33) dimana untuk *cluster* kedua ini mempunyai nilai SSE sebesar 32,1829.

3. *Cluster* 3 memiliki anggota berjumlah 2 provinsi yaitu Jawa Barat (12) dan Jawa Timur (15) dengan nilai SSE sebesar 45,5284.

Dari tiga pengelompokkan yang terbentuk, dapat dilihat bahwa *cluster* 2 memiliki nilai SSE yang paling kecil dimana hasil nilai SSE yaitu sebesar 32,1829 dengan 4 anggota. Sedangkan untuk *cluster* yang mempunyai nilai SSE dengan kemiripan karakteristik terjauh dengan objek yang lain yaitu pada *cluster* 1 sebesar 87,9631.

4.7. Pemilihan Metode Terbaik

Dalam mengetahui kinerja dari suatu metode *clustering*, kriteria yang digunakan adalah dua nilai dari simpangan baku yaitu simpangan baku dalam kelompok (S_w) serta simpangan baku antar kelompok (S_b). Nilai simpangan baku dalam kelompok yang semakin kecil serta simpangan baku antar kelompok yang semakin besar maka semakin baik pula kinerja dari metode yang digunakan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.7.1. Simpangan Baku pada Metode *Complete Linkage* dan *Average Linkage*

Berdasarkan pada Lampiran F dan G menunjukkan bahwa hasil uji coba 2 sampai 6 pada metode *Complete Linkage* dan *Average Linkage* menunjukkan hasil yang sama. Berikut ini adalah salah satu contoh dalam perhitungan simpangan baku dalam kelompok (S_w) serta simpangan baku antar kelompok (S_b) jika akan membentuk 3 *cluster*.

1. Simpangan baku dalam kelompok (S_w)

   Sebelum melakukan perhitungan nilai S_w, yang harus dilakukan adalah menghitung nilai simpangan baku kelompok ke-k (S_k). Berikut ini merupakan contoh dalam menghitung nilai S_k:

   $S_1 = \sqrt{\frac{(x_1-\bar{x_1})^2+(x_2-\bar{x_1})^2+...+(x_{30}-\bar{x}_1)^2}{n-1}}$

   $S_1 = \sqrt{\frac{(0{,}3686-(-0{,}2219))^2+(0{,}1112-(-0{,}2219))^2+...+(-0{,}4409-(-0{,}2219))^2}{30-1}}$

   $S_1 = 1,5626$

   $S_2 = \sqrt{\frac{(x_1-\bar{x_2})^2+(x_2-\bar{x_2})^2+(x_3-\bar{x}_2)^2}{n-1}}$

   $S_2 = \sqrt{\frac{1{,}9629-(1{,}8648))^2+(1{,}6556-(1{,}8648))^2+(1{,}9760-(1{,}8648))^2}{3-1}}$

   $S_2 = 0,24409$

   Untuk *cluster* ketiga tidak memiliki nilai simpangan baku dalam kelompok serta nilai simpangan baku tidak berpengaruh karena hanya beranggotakan satu provinsi saja.

   Setelah memperoleh nilai simpangan baku tiap kelompok maka selanjutnya dapat menghitung nilai simpangan baku dalam *cluster* (S_w) seperti berikut ini:

   $S_w = \frac{1}{K}\Sigma_{k=1}^{K} S_k$

   $S_w = \frac{1}{3}(1,5626 + 0,24409)$

   $S_w = 0,6023$

   Jadi nilai simpangan baku dalam *cluster* (S_w) sebesar 0,6023 artinya hal ini

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

menunjukkkan bahwa kelompok anggota dalam satu *cluster* memiliki tingkat homogenitas yang cukup tinggi. Semakin kecil nilai simpangan baku dalam *cluster* menunjukkan bahwa dalam suatu *cluster* memiliki homogenitas yang tinggi.

2. Simpangan baku antar kelompok (S_b)

   Dalam menghitung nilai simpangan baku antar kelompok (S_b) membutuhkan nilai rata-rata setiap *cluster* yang tertera pada Lampiran I. Sebelum menghitung nilai S_b terlebih dahulu menghitung nilai $\bar{X}$ seperti berikut ini:

   $\bar{X} = \frac{-0{,}2219+1{,}8648+1{,}0630}{3}$

   $\bar{X} = 0,9020$

   Selanjutnya melakukan perhitungan nilai S_b seperti di bawah ini :

   $S_b = \left[\frac{1}{K-1}\Sigma_{k=1}^{K}(\bar{X}_k - \bar{X})^2\right]^{\frac{1}{2}}$

   $S_b = \left[\frac{(\bar{X}_1-\bar{X})^2+(\bar{X}_2-\bar{X})^2+(\bar{X}_3-\bar{X})^2}{K-1}\right]^{\frac{1}{2}}$

   $S_b = \left[\frac{(-0{,}2219-0{,}9020)^2+(1{,}8648-0{,}9020)^2+(1{,}0630-0{,}9020)^2}{3-1}\right]^{\frac{1}{2}}$

   $S_b = 1,4843$

   Berdasarkan perhitungan diatas diperoleh nilai simpangan baku antar *cluster* S_b sebesar 1,4843. Semakin tinggi nilai S_b menunjukkan bahwa semakin lebar rentang variansi antar *cluster* artinya antar *cluster* satu dengan *cluster* yang lainnya memiliki heterogenitas yang cukup tinggi.

3. Rasio Simpangan Baku

   Setelah memperoleh hasil perhitungan nilai (S_w) dan S_b, selanjutnya adalah menghitung nilai rasio (S_w) terhadap S_b sebagai berikut:

   $S = \frac{S_w}{S_b}$

   $S = \frac{0{,}6023}{1{,}4843}$

   $S = 0,4057$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Jadi nilai rasio simpangan baku dalam *cluster* terhadap simpangan baku antar *cluster* menggunakan metode *Complete Linkage* dan *Average Linkage* adalah 0,4057. Nilai rasio simpangan baku yang semakin kecil menunjukkan bahwa semakin baik kinerja dari metode yang digunakan.

4.7.2. Simpangan Baku pada Metode *Ward*

1. Simpangan baku dalam kelompok (S_w)

   Sebelum melakukan perhitungan nilai S_w terlebih dahulu yang harus dilakukan adalah menghitung nilai simpangan baku kelompok ke-k (S_k). Dibawah ini merupakan contoh dalam menghitung nilai S_k:

   $S_1 = \sqrt{\frac{(x_1-\bar{x_1})^2+(x_2-\bar{x_1})^2+...+(x_4-\bar{x}_1)^2}{n-1}}$

   $S_1 = \sqrt{\frac{(0{,}36855-(0{,}9507))^2+(0{,}7158-(0{,}9507))^2+...+(1{,}0630-(0{,}9507))^2}{4-1}}$

   $S_1 = 0,9461$

   $S_2 = \sqrt{\frac{(x_1-\bar{x_2})^2+(x_2-\bar{x_2})^2+...+(x_{28}-\bar{x}_2)^2}{n-1}}$

   $S_2 = \sqrt{\frac{0{,}1112-(-0{,}2765))^2+(-0{,}1804-(-0{,}2765))^2+(-0{,}4409-(-0{,}2765))^2}{38-1}}$

   $S_2 = 1,0727$

   $S_3 = \sqrt{\frac{(x_1-\bar{x_2})^2+(x_2-\bar{x_2})^2}{n-1}}$

   $S_3 = \sqrt{\frac{1{,}9629-(1{,}9695))^2+(1{,}9760-(1{,}9695))^2}{2-1}}$

   $S_3 = 0,0092$

   Setelah memperoleh nilai simpangan baku tiap kelompok maka selanjutnya dapat menghitung nilai simpangan baku dalam *cluster* (S_w) seperti dibawah ini:

   $S_w = \frac{1}{K}\Sigma_{k=1}^{K} S_k$

   $S_w = \frac{1}{3}(0,9461 + 1,0727 + 0,0092)$

   $S_w = 0,6760$

   Nilai simpangan baku dalam *cluster* (S_w) menggunakan metode *Ward* sebesar 0,6760. Hal ini menunjukkkan bahwa kelompok anggota dalam satu *cluster*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

memiliki tingkat homogenitas yang cukup tinggi. Semakin kecil nilai simpangan baku dalam *cluster* menunjukkan bahwa dalam suatu *cluster* memiliki homogenitas yang tinggi.

2. Simpangan baku antar kelompok (S_b)

   Dalam menghitung nilai simpangan baku antar kelompok (S_b) membutuhkan nilai rata-rata setiap *cluster* yang tertera pada Lampiran J. Sebelum menghitung nilai S_b terlebih dahulu menghitung nilai $\bar{X}$ seperti berikut ini:

   $\bar{X} = \frac{0{,}9507+(-0{,}2765)+1{,}9695}{3}$

   $\bar{X} = 0,8812$

   Selanjutnya melakukan perhitungan nilai S_b seperti di bawah ini :

   $S_b = \left[\frac{1}{K-1}\Sigma_{k=1}^{K}(\bar{X}_k - \bar{X})^2\right]^{\frac{1}{2}}$

   $S_b = \left[\frac{(\bar{X}_1-\bar{X})^2+(\bar{X}_2-\bar{X})^2+(\bar{X}_3-\bar{X})^2}{K-1}\right]^{\frac{1}{2}}$

   $S_b = \left[\frac{(0{,}9507-0{,}8812)^2+(-0{,}2765-0{,}8812)^2+(1{,}9695-0{,}8812)^2}{3-1}\right]^{\frac{1}{2}}$

   $S_b = 1,3919$

   Berdasarkan perhitungan diatas diperoleh nilai simpangan baku antar *cluster* S_b sebesar 1,3919. Semakin tinggi nilai S_b menunjukkan bahwa semakin lebar rentang variansi antar textitcluster artinya antar *cluster* satu dengan *cluster* yang lainnya memiliki heterogenitas yang cukup tinggi.

3. Rasio Simpangan Baku

   Setelah memperoleh hasil perhitungan nilai (S_w) dan S_b, selanjutnya adalah menghitung nilai rasio (S_w) terhadap S_b sebagai berikut:

   $S = \frac{S_w}{S_b}$

   $S = \frac{0{,}6760}{1{,}3919}$

   $S = 0,4857$

   Jadi nilai rasio simpangan baku dalam *cluster* terhadap simpangan baku antar

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

cluster menggunakan metode *Ward* adalah 0,4857. Semakin kecil nilai rasio simpangan baku dalam *cluster* terhadap simpangan baku antar *cluster* menunjukkan bahwa semakin baik kinerja dari metode yang digunakan.

4.7.3. Simpangan Baku Keseluruhan *Cluser*

Hasil rasio simpangan baku dalam *cluster* (S_w), rasio simpangan baku antar *cluster* (S_b), dan rasio simpangan baku dalam *cluster* terhadap simpangan baku antar *cluster* (S) metode *Complete Linkage* dan *Average Linkage* tertera pada Tabel 4.3.

Tabel 4.3 Perbandingan Rasio Simpangan Baku *Complete Linkage* dan *Average Linkage*

Jumlah *Cluster*	S_w	S_b	S
2 *Cluster*	0,92688	1,44625	0,64089
3 *Cluster*	0,60225	1,48432	0,40573
4 *Cluster*	0,39295	1,67645	0,23440
5 *Cluster*	0,29093	1,81692	0,16012
6 *Cluster*	0,24090	2,03703	0,11826

Hasil dari rasio simpangan baku dalam *cluster* (S_w), rasio simpangan baku antar *cluster* (S_b), dan rasio simpangan baku dalam *cluster* terhadap simpangan baku antar *cluster*(S) metode *Ward* tertera pada Tabel 4.4.

Tabel 4.4 Perbandingan Rasio Simpangan Baku *Ward*

Jumlah *Cluster*	S_w	S_b	S
2 *Cluster*	1,28567	1,10789	1,16046
3 *Cluster*	0,67600	1,39186	0,485682
4 *Cluster*	0,39268	1,59351	0,24642

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Jumlah *Cluster*	S_w	S_b	S
5 *Cluster*	0,35376	1,69735	0,20847
6 *Cluster*	0,18405	2,12284	0,08670

Berdasarkan pada Tabel 4.3 dan 4.4 hasil dari rasio simpangan baku (S) menunjukkan bahwa metode *Ward* dengan 6 *cluster* mempunyai kinerja yang paling baik dalam membentuk *cluster* provinsi berdasarkan variabel jenis penyakit diantara dua metode yang lainnya yaitu metode *Complete Linkage* dan *Average Linkage*. Hal tersebut dapat diketahui dari nilai simpangan baku dalam *cluster* (S_w) terhadap simpangan baku antar *cluster* (S_b) pada metode *Ward* dengan 6 *cluster* memiliki nilai yang paling minimum diantara ketiga metode tersebut. Dari hasil perbandingan rasio simpangan baku menunjukkan bahwasannya 6 *cluster* dari metode *Ward* yang memiliki kinerja terbaik maka 6 *cluster* provinsi yang terbentuk adalah tertera pada Tabel 4.5.

Tabel 4.5 Jumlah *Cluster* dan Anggotanya Metode Ward

Provinsi	***Cluster***
Aceh	I
Sumatra Utara,Sumatra Barat,Riau Jambi, Sumatra Selatan, Bengkulu, Lampung Kep. Bangka Belitung, Kep. Riau, DI Yogyakarta Banten, Bali, NTB, NTT, Kalimantan Barat Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur	II
Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah Sulawesi Tenggara, Gorontalo, Sulawesi Barat Maluku, Maluku Utara, Papua Barat	II
DKI Jakarta	III

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Provinsi	*Cluster*
Jawa Tengah	IV
Papua	V
Jawa Barat, Jawa Timur	VI

Berdasarkan hasil pengelompokan yang telah terbentuk, didapatkan 6 *cluster* provinsi yang digambarkan melalui peta Indonesia yang dapat dilihat pada Gambar 4.5 sebagai berikut :

Gambar 4.5 Ilustrasi Peta Hasil *Cluster*

Setelah *cluster* terbentuk tahap selanjutnya adalah memberikan spesifikasi atau ciri dalam menggambarkan isi *cluster* tersebut atau disebut dengan intepretasi *cluster* sebagai berikut :

1. *Cluster* pertama memiliki anggota berjumlah 1 provinsi dengan nilai rata-rata kasus penyakit sebesar 1.643 kasus, dimana untuk *cluster* pertama ini memiliki nilai rata-rata kasus penyakit paling rendah dari *cluster* yang lain. Oleh karena itu untuk *cluster* kedua ini dapat dikelompokkan sebagai kelompok provinsi dengan tingkat kesehatan yang sangat baik jika ditinjau dari kasus penyakit

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

yang ada. Untuk *cluster* pertama ini penyakit yang sering terjadi adalah Campak.

2. *Cluster* kedua terdiri dari 28 provinsi dimana untuk nilai rata-rata kasus penyakit sebesar 2.068 kasus. Untuk *cluster* 2 dapat dikelompokkan sebagai kelompok provinsi dengan tingkat kesehatan yang baik bila ditinjau dari kasus penyakit yang ada. Penyakit yang sering terjadi pada *cluster* adalah Difteri.

3. *Cluster* ketiga terdiri dari 1 provinsi dengan nilai rata-rata kasus penyakit sebesar 10.014 kasus. Untuk *cluster* ketiga ini dapat dikelompokkan sebagai kelompok provinsi dengan tingkat kesehatan yang cukup baik bila ditinjau dari kasus penyakit yang ada. Pada *cluster* penyakit yang sering terjadi adalah penyakit HIV.

4. *Cluster* keempat terdiri dari 1 provinsi saja dengan nilai rata-rata kasus penyakit sebesar 14.669. Apabila diurutkan berdasarkan nilai rata-rata kasus penyakitnya untuk untuk *cluster* 4 ini dapat dikelompokkan sebagai kelompok provinsi dengan tingkat kesehatan yang kurang baik. Penyakit yang sering terjadi pada *cluster* keempat adalah AIDS.

5. *Cluster* kelima memiliki anggota berjumlah 1 provinsi dengan nilai rata-rata kasus penyakit sebesar 17.295 kasus. Untuk *cluster* kelima dapat dikelompokkan sebagai kelompok provinsi dengan tingkat kesehatan yang tidak baik bila ditinjau dari kasus penyakit yang ada. Penyakit yang sering terjadi pada *cluster* ini adalah Malaria.

6. *Cluster* keenam terdiri dari 2 provinsi dengan nilai rata-rata kasus penyakit sebesar 22.904 kasus. Karena *cluster* ini memiliki nilai rata-rata kasus penyakit paling tinggi dari 6 *cluster* yang ada. Untuk *cluster* 6 ini dapat dikelompokkan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

sebagai kelompok provinsi dengan tingkat kesehatan yang sangat tidak baik bila ditinjau dari kasus penyakit yang ada. Penyakit yang sering terjadi pada *cluster* ketiga ini adalah TBC, DBD, Kusta dan Pneumonia.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB V

PENUTUP

Pada bab ini akan diberikan simpulan dan saran-saran yang dapat diambil berdasarkan materi-materi yang telah dibahas pada bab-bab sebelumnya.

5.1. Simpulan

Berdasarkan pada hasil pembahasan mengenai data jumlah kasus penyakit di Indonesia menggunakan ketiga metode yaitu *Complete Linkage*, *Average Linkage*, dan *Ward* dapat ditarik kesimpulan sebagai berikut :

1. Apabila ditinjau berdasarkan nilai simpangan baku dalam kelompok (S_w) terhadap simpangan baku antar kelompok (S_b) :

   Hasil rasio simpangan baku paling kecil terdapat pada metode *Ward* dengan 6 *cluster* yaitu sebesar 0,08670. Oleh karena itu metode *Ward* merupakan metode yang paling baik jika dibandingkan dengan metode *Complete Linkage* dan *Average Linkage* karena pada metode *Ward* memiliki nilai rasio simpangan baku terkecil.

2. Hasil dari uji coba 2 hingga 6 *cluster* yang terbentuk berdasarkan metode *Complete Linkage*, *Average Linkage*, dan *Ward*, menunjukkan bahwa yang terbaik adalah dengan membentuk 6 *cluster* yakni untuk hasil *cluster* dengan metode *Ward* didapatkan hasil *cluster* pertama terdiri dari 1 provinsi, *cluster* kedua terdiri dari 28 provinsi, *cluster* ketiga, *cluster* keempat, dan *cluster* kelima, masing-masing terdiri dari 1 provinsi, serta *cluster* keenam terdiri dari

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2 provinsi. Urutan *cluster* provinsi berdasarkan kelompok yang telah terbentuk diperoleh bahwa untuk tingkat kesehatan sangat baik, baik, cukup baik, kurang baik, tidak baik, dan sangat tidak baik berturut-turut yaitu *cluster* 1, *cluster* 2, *cluster* 3, *cluster* 4, *cluster* 5, dan *cluster* 6.

5.2. Saran

Setelah membahas mengenai analisis *cluster* dalam pengelompokkan provinsi di Indonesia berdasarkan variabel jenis penyakit menggunakan metode *Complete Linkage*, *Average Linkage*, dan *Ward*, penulis ingin menyampaikan beberapa saran yaitu mengingat bahwasannya ruang lingkup metode *cluster* yang cukup banyak. Bagi peneliti yang juga ingin melakukan penelitian mengenai *clustering* dapat menggunakan metode hierarki yang lainnya seperti metode *Single Linkage* dan Centroid, serta dapat menggunakan metode non hierarki ataupun dapat menggunakan metode *Fuzzy C-Means*, *K-Means*, *K-Modes* dll.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR PUSTAKA

Azizah, I. A. (2011). *Gambaran Prevelensi dan Resiko Penyakit Kardiovaskular pada Penduduk Dewasa di Provinsi Jakarta*. Jakarta: Universitas Indonesia.

Barakbah, A.R., & Korei, Arai. (2004). *Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering*. *IES*, 409-413.

Fadlina, A. (2015). *Penerapan Metode Agglomerative Hierarchical Clustering Untuk Klasifikasi Kabupaten/ Kota di Provinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana*. Malang: UIN Maulana Malik Ibrahim.

Febrina, T. L. (2014). *Perbandingan Kinerja Metode Complete Linkage, Metode Average Linkage, dan Metode K-Means Dalam Menentukan Hasil Analisis Cluster*. Yogyakarta: Universitas Negeri Yogyakarta.

Hartini, E. (2014). *Metode Clustering Hirarki*. Pusat Pengembangan Teknologi Informasi dan Komputasi BATAN.

Fauzia, Y. W., & Paramita, A. (2016). *Pelayanan Kesehatan Mental dalam Hubungannya dengan Disibilitas dan Gaya Hidup Masyarakat Indonesia (Analisis Lajut Riskesdas 2007 dan 2013)*. *Buletin Penelitian Sistem Kesehatan*, 19, 1, 99-107.

Fitriati, U. (2017). *Pengaruh Pendidikan Kesehatan Tentang Cara Penularan HIV dan AIDS Terhadap Persepsi pada Odha di Kelurahan Sosromenduran Kota Yogyakarta*. Yogyakarta: Universitas Aisyiyah Yogyakarta.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Halim, R. M. G. (2016).*Campak pada Anak*.

Harli, N. F. (2014). *Penatalaksanaan Difteri*. *J Indon Med Assoc*, 64, 12, 541-545.

Ika, A., A., D.(2014). *Pneomonia pada Balita di Indonesia*.

Ilo,& Who.(2018). *Pedoman Bersama ILO/WHO Tentang Pelayanan Kesehatan dan HIV/AIDS*. Jakarta : Direktorat Pengawasan Kesehatan Kerja.

Indah, M.(2018). *Dicari Para Pemimpin Untuk Dunia Bebas TB*.

Johnson, Richard.A. Wichern, Dean.W.(1996).*Applied Multivariate Stastistical Analysis*. New Delhi: Prentice-Hall.

Kemenkes, RI. (2017). *Situasi Penyakit Demam Berdarah di Indonesia Tahun 2017*. Jakarta: Kementerian Kesehatan RI Pusat Data dan Informasi.

Kemenkes, RI.(2017). *Difteri Menular, Berbahaya dan Mematikan, Namun Bisa Dicegah Dengan Imunisasi*.

Kemenkes,RI. (2017). *Hapuskan Stigma dan Deskriminasi terhadap Kusta*

Kemenkes, RI. (2018). *Potret Sehat Indonesia dari Riskesdas 2018*.

Latan, H.(2014). *Aplikasi analisis data statistik untuk ilmu sosial sains dengan STATA*. Bandung : Alfabeta.

Laeli, S.(2014). *Analisis Cluster dengan Average Linkage Method dan Wards Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link*. Yogyakarta : Universitas Negeri Yogyakarta.

Marice, S., Aprildah, N. S. (2014). *Faktor Risiko Tumor Payudara Pada Perempuan Umur 225-65 Tahun di Lima Kelurahan Kecamatan Bogor Tengah*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Ningsih, S., Wahyuningsih, S., & Nasution, N.Y. (2016). *Perbandingan Kinerja Metode Complete Linkage dan Average Linkage dalam Menentukan Hasil Analisis Cluster (Studi Kasus Produksi Palawija Provinsi Kalimantan Timur* 2014/2015*). Prosiding SeminarSains dan Teknologi FMIPA Unmul*, 01, 01.

Ramadhani, L., Purnamasari, I.(2018). *Penerapan Metode Complete Linkage dan MetodeHierarchical Clustering Multiscale Bootstrap(Studi Kasus: Kemiskinan Di Kalimantan Timur Tahun 2016). Jurnal Eksponensial*, 09,01.

Rencher, A. (2002). *Method of Multivariate Analysis. 2nd ed.*. John Wiley and Sons, Inc. New York.

Satria, F., Aziz A. RZ. (2016). *Perbandingan Kinerja Metode Ward dan K-Means dalam Menentukan Cluster Data Mahasiswa Permohonan Beasiswa (STUDI KASUS : STMIK PRINGSEWU). Jurnal TIM Darmajaya*, 2, 1.

Sholiha, A. (2015).*Perbandingan Analisis Klaster Menggunakan Metode Single Linkage , Complete Linkage , Average Linkage Dan K-Means Untuk Pengelompokan Jenis Ternak di Kabupaten Semarang*. Semarang : Universitas Negeri Semarang.

Simarora, B. (2005). *Analisis Multivariat Pemasaran Edisi Pertama*. Jakarta : PT. Gramedia Pustaka tama.

Soraya, Yani. (2011).*Perbandingan Kinerja Metode Single Linkage, Metode Complete Linkage dan Metode K-Means dalam Analisis Cluster*. Semarang : Universitas Negeri Semarang.

Zein, U. (2017), *Daerah Endemik Malaria di Indonesia, Fokus pada Malaria Ovale*. *Jurnal Kedokteran Methodist*, 8, 8, 613-617.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id