

**ANALISIS *CLUSTER HIRARKI* METODE *AVERAGE LINKAGE*  
BERDASARKAN JUMLAH KRIMINALITAS  
DI INDONESIA TAHUN 2019**

**SKRIPSI**

**OLEH  
RENTA WINDASARI  
NIM. 16610058**



**JURUSAN MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2020**

**ANALISIS *CLUSTER HIRARKI* METODE *AVERAGE LINKAGE*  
BERDASARKAN JUMLAH KRIMINALITAS  
DI INDONESIA TAHUN 2019**

SKRIPSI

Oleh  
**RENTA WINDASARI**  
NIM. 16610058

Telah Diperiksa dan Disetujui untuk Diuji  
Tanggal: 14 Mei 2020

Pembimbing I,



Ari Kusumastuti, M.Si., M.Pd  
NIP. 197705212005012004

Pembimbing II,



Muhammad Khudzaifah, M.Si  
NIDT. 19900511201608011057

Mengetahui,

Ketua Jurusan Matematika



Dr. Usman Pagalay, M.Si  
NIP. 196504142003121001

**ANALISIS *CLUSTER HIRARKI* METODE *AVERAGE LINKAGE*  
BERDASARKAN JUMLAH KRIMINALITAS  
DI INDONESIA TAHUN 2019**

SKRIPSI

Oleh  
**RENTA WINDASARI**  
NIM. 16610058

Telah dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima sebagai salah satu Persyaratan  
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal 14 Mei 2020

Penguji Utama : Dr. H. Imam Sujarwo, M.Pd  
Ketua Penguji : Angga Dwi Mulyanto M,Si  
Sekretaris Penguji : Ari Kusumastuti, M.Si., M.Pd  
Anggota Penguji : Muhammad Khudzaifah, M.Si



Mengetahui

Ketua Jurusan Matematika



Dr. Usman Pagalay, M.Si  
NIP. 196504142003121001

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Renta Windasari

NIM : 16610058

Jurusan : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Analisis *Cluster* Hirarki Metode *Average Linkage* Berdasarkan Jumlah Kriminalitas di Indonesia Tahun 2019

menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar rujukan. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 14 Mei 2020

Yang membuat pernyataan,



Renta Windasari  
NIM. 16610058

## MOTTO

*“Angin tidak berhembus untuk menggoyangkan pepohonan, melainkan menguji kekuatan akarnya”*



## PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillah Robbil`alamin, dengan mengucap syukur kepada Allah Swt, skripsi ini penulis persembahkan kepada:

Ayahanda Mochamad Rochim dan Ibunda Jiati tercinta.

Saudara-saudara dan teman-teman yang senantiasa memberikan semangat, arahan, dan juga materi, sehingga penulis dapat berjuang menyelesaikan perkuliahan di Universitas Islam Negeri Maulana Malik Ibrahim Malang.



## KATA PENGANTAR

*Assalamu`alaikum Wr. Wb.*

Puji syukur *Alhamdulillah* penulis haturkan ke hadirat Allah Swt yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan skripsi ini dengan baik.

Skripsi yang berjudul “Analisis *Cluster* Hirarki Metode *Average Linkage* Berdasarkan Jumlah Kriminalitas di Indonesia Tahun 2019” ini penulis susun untuk memenuhi salah satu syarat menyelesaikan pendidikan program studi strata satu (S-1) di Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Penulis haturkan ucapan terima kasih seiring do`a dan harapan *jazakumullah ahsanal jaza`* kepada semua pihak yang telah membantu menyelesaikan skripsi ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Prof. Dr. Abd. Haris, M.Ag, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah banyak memberikan pengetahuan dan pengalaman yang berharga.
2. Dr. Sri Harini, M.Si, selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang
3. Dr. Usman Pagalay, M.Si, selaku Ketua Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang, yang telah banyak memberikan motivasi belajar kepada penulis.
4. Ari Kusumastuti, M.Si., M.Pd, selaku dosen pembimbing I yang selalu sabar membimbing dengan segala ilmu yang dimiliki kepada penulis.
5. Hawzah Sa`adati, M.Si, selaku dosen pembimbing II yang telah memberikan bimbingan, arahan, dan berbagai ilmunya kepada penulis.

6. Ach. Nashichuddin, M.A, selaku dosen wali yang selalu memberikan motivasi dan arahan kepada penulis.
7. Segenap sivitas akademika Jurusan Matematika, terutama seluruh dosen, terima kasih atas segenap ilmu dan bimbingannya selama masa perkuliahan.
8. Kedua orang tua yang tak punya ruang jeda untuk berdoa serta bekerja bersimbah pelu demi pendidikan, kebahagiaan, dan kesuksesan penulis.
9. Sahabat-sahabat seperjuangan, mahasiswa Matematika angkatan 2016, terima kasih atas segala pengalaman berharga dan kenangan yang terukir rapid an abadi.
10. Sahabat-sahabat “Musyrif/Musyrifah” Ma`had Sunan Ampel Al-Aly Universitas Islam Negeri Maulana Malik Ibrahim Malang, terkhusus Musyrif/Musyrifah “*Miracle*” serta para Musyrifah Mabna Ummu Salamah 2017-2019 dan para Musyrifah Mabna Fatimah Az-Zahra 2019-2020, terima kasih atas segala pengalaman serta pengabdian yang sangat berharga di Ma`had.
11. Semua pihak yang secara langsung maupun tidak langsung yang tidak mungkin penulis sebut satu persatu, terima kasih atas keikhlasan bantuan moral, material dan spiritual yang sudah diberikan kepada penulis.

Penulis berharap semoga skripsi ini dapat memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. *Aamiin Ya Rabbal Aalamiin.*

*Wassalamu`alaikum Wr. Wb.*

Malang, 14 Mei 2020

Penulis

## DAFTAR ISI

**HALAMAN JUDUL**

**HALAMAN PENGAJUAN**

**HALAMAN PENGESAHAN**

**HALAMAN PERNYATAAN KEASLIAN TULISAN**

**HALAMAN MOTTO**

**HALAMAN PERSEMBAHAN**

**KATA PENGANTAR..... vi**

**DAFTAR ISI..... viii**

**DAFTAR TABEL ..... x**

**DAFTAR GAMBAR..... xi**

**ABSTRAK ..... xii**

**ABSTRACT ..... xiii**

**مستخلص البحث..... xiv**

### **BAB 1 PENDAHULUAN**

1.1	Latar Belakang .....	1
1.2	Rumusan Masalah .....	3
1.3	Tujuan Penelitian.....	4
1.4	Manfaat Penelitian.....	4
1.5	Batasan Masalah.....	5
1.6	Sistematika Penulisan.....	5

### **BAB II KAJIAN PUSTAKA**

2.1	Analisis <i>Cluster</i> .....	7
2.2	Prosedur Analisis <i>Cluster</i> .....	8
2.3	Metode Pengelompokan .....	9
2.3.1	Metode Hirarki (Hierarchical Method).....	9
2.3.2	Metode Non-Hirarki (Non Hierarchical Method) .....	14
2.3.3	Metode Two Step <i>Cluster</i> .....	15

2.4	Metode Menghitung Jarak.....	15
2.5	Uji Asumsi Analisis <i>Cluster</i> .....	19
2.6	Proses Dasar Analisis <i>Cluster</i> .....	21
	2.6.1 Mengukur Kesamaan Antar Objek ( <i>Similarity</i> ) .....	22
	2.6.2 <i>Average Linkage</i> .....	22
2.7	Kajian Jenis Kriminalitas di Indonesia.....	23
2.8	Analisis <i>cluster</i> dalam Pandangan Islam .....	23

### BAB III METODOLOGI PENELITIAN

3.1	Pendekatan Penelitian.....	25
3.2	Sumber Data .....	27
3.3	Variabel Penelitian .....	28
3.4	Analisis Data .....	28
	3.4.1 Pengelompokan Jumlah Kriminalitas di Indonesia .....	28
	3.4.2 Karakteristik Setiap Kelompok.....	30
3.5	Flowchart Analisis Cluster .....	31

### BAB IV PEMBAHASAN

4.1	Analisis <i>cluster</i> pada Jumlah Kriminalitas .....	33
	4.1.1 Uji Asumsi Analisis <i>Cluster</i> .....	33
4.2	Penerapan Analisis <i>cluster</i> pada Jumlah Kriminalitas .....	35
	4.2.1 Jarak Manhattan.....	36
	4.2.2 Metode <i>Average Linkage</i> .....	38
	4.2.2 Menentukan Jumlah Anggota <i>Cluster</i> .....	45
4.3	Karakteristik Setiap Kelompok .....	46
	4.3.1 Uji Validitas dan Profiling <i>Cluster</i> .....	47

### BAB V PENUTUP

5.1	Kesimpulan.....	49
5.2	Saran .....	49

### DAFTAR RUJUKAN

### LAMPIRAN-LAMPIRAN

### RIWAYAT HIDUP

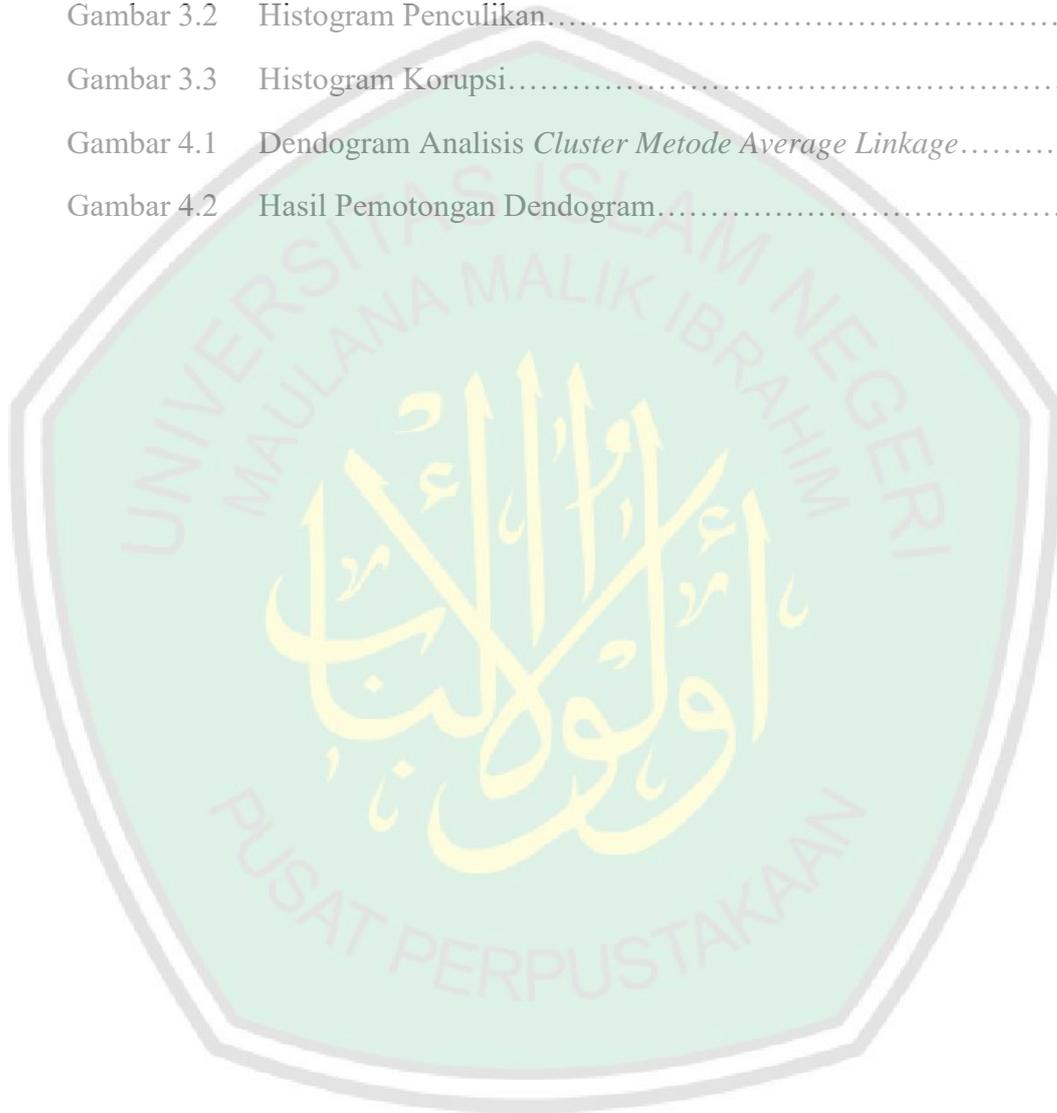
### BUKTI KONSULTASI

## DAFTAR TABEL

Tabel 2.1	Karakteristik Nilai KMO.....	20
Tabel 3.1	Deskripsi Data.....	25
Tabel 3.2	Data.....	27
Tabel 3.2	Keterangan Simbol <i>Flowchart</i> .....	32
Tabel 4.1	KMO dan <i>Bartlett`s Test</i> .....	33
Tabel 4.2	Uji VIF.....	34
Tabel 4.3	Uji VIF.....	34
Tabel 4.4	Uji VIF.....	34
Tabel 4.5	Matriks Jarak Manhatt.....	37
Tabel 4.6	Matriks Jarak Baru.....	39
Tabel 4.7	<i>Cluster Joined</i> .....	39
Tabel 4.8	Hasil Analisis <i>Cluster</i> .....	43
Tabel 4.9	Anggota Masing-Masing <i>Cluster</i> .....	45
Tabel 4.10	Perbedaan Karakteristik Setiap Kelompo.....	46
Tabel 4.11	<i>Cluster 1</i> .....	47
Tabel 4.12	<i>Cluster 2</i> .....	47

## DAFTAR GAMBAR

Gambar 2.1	Contoh Dendogram <i>Average Linkage</i> .....	10
Gambar 3.1	Histogram Mempekerjakan anak dibawah umur.....	26
Gambar 3.2	Histogram Penculikan.....	26
Gambar 3.3	Histogram Korupsi.....	26
Gambar 4.1	Dendogram Analisis <i>Cluster Metode Average Linkage</i> .....	40
Gambar 4.2	Hasil Pemotongan Dendogram.....	45



## ABSTRAK

Windasari, Renta. 2020. Analisis *Cluster* Hirarki metode *Average Linkage* Berdasarkan Jumlah Kriminalitas di Indonesia Tahun 2019. Skripsi. Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Ari Kusumastuti, M.Si., M.Pd. (II) Muhammad Khudzaifah M.Si.

Kata Kunci: *cluster, hirarki, average linkage, jarak manhattan, kriminalitas*

Kriminalitas merupakan segala sesuatu yang melanggar hukum atau tindak kejahatan. Terdapat berbagai jenis kriminalitas seperti pencurian, perampokan, korupsi dan lain sebagainya. Salah satu metode yang dapat digunakan untuk menganalisis data tindak kriminalitas yaitu metode analisis *cluster* yang merupakan salah satu teknik analisis multivariat yang bertujuan untuk mengklasterkan data observasi ataupun variabel-variabel ke dalam kelompok yang memiliki kedekatan jarak dan tingkat kemiripan yang tinggi. Tujuan dari pengklasteran tersebut yaitu untuk mengetahui daerah-daerah yang memiliki kedekatan atau kemiripan jenis kriminalitas serta untuk mengurangi jumlah kriminalitas di Indonesia yang sering terjadi. Dalam analisis *cluster* terdapat tiga metode, yaitu metode hirarki, *non* hirarki, dan *two step cluster*. Metode hirarki terbagi atas 5 metode antara lain metode *Complete Linkage, Average Linkage, Single Linkage, Centroid, dan Ward Linkage*. Pada penelitian ini digunakan Metode *Average Linkage*. Metode *Average Linkage* menggunakan nilai rata-rata pada variabel kemudian diterapkan dalam pengelompokan jumlah kriminalitas di Indonesia Tahun 2019. Hasil pengelompokan pada penelitian ini terdapat 2 *cluster* yaitu *cluster* 1 terdiri dari Aceh, Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, Kepulauan Riau, Metro Jaya, Jawa Barat, DI Yogyakarta, Bali, Nusa Tenggara Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Sumatera Utara, Sumatera Barat, Sumatera Selatan, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Barat, Maluku, Papua Barat, Papua dan *cluster* 2 terdiri dari Sulawesi Selatan saja.

## ABSTRACT

Windasari, Renta. 2020. **Cluster Hierarchy Analysis of the Average Linkage Method Based on the Number of Crimes in Indonesia 2019**. Thesis. Mathematics Department, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Advisor: (I) Ari Kusumastuti, M.Si., M.Pd. (II) Muhammad Khudzaifah, M.Si.

Keywords: *cluster*, hierarchy, average linkage, manhattan distance, crime

Crime is anything that againsts the law or a crime. There are various types of crime such as theft, robbery, corruption and so on. One of method that can be used to analyze criminal data is cluster analysis method, which is a multivariate analysis technique that aims to cluster observational data or variables into groups that have close distance and a high degree of similarity. The aim of the clustering is to find out locations that have a close or similar type of crime and to reduce the amount of crime that often occurs in Indonesia. There are three methods in cluster analysis, namely the hierarchical method, non-hierarchical, and two step cluster methods. The hierarchy method is divided into 5 methods including Complete Linkage, Average Linkage, Single Linkage, Centroid, and Ward Linkage. This research uses the Average Linkage method. The Average Linkage method uses the average value of the variables then applied in grouping the amount of crime in Indonesia in 2019. There are 2 clusters as the result of this research namely cluster 1 consisting of Aceh, Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, Riau Islands, Metro Jaya, West Java, DI Yogyakarta, Bali, West Nusa Tenggara, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, North Maluku, North Sumatra, West Sumatra, South Sumatra, Central Java, East Java, Banten, East Nusa Tenggara, West Kalimantan, Maluku, West Papua, Papua and cluster 2 comprise only South Sulawesi.

## مستخلص البحث

وينداساري. رنتا. ٢٠٢٠. تحليل "مجموعة" بالطريقة "Average Linkage" بناء على عدد الجريمة في إندونيسيا عام ٢٠١٩. البحث الجامعي. قسم الرياضيات. كلية العلوم والتكنولوجيا. الجامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج. المشرف: ١. أري كوسوماستوتي الماجستير، ٢.

هوزة سعادتي الماجستير

الكلمة الأساسية: مجموعة (Cluster)، تسلسل، Average Linkage، مسافة "manhattan"، الجريمة

الجريمة هي الأساس الذي خارج على القانون أو الإجرام. في العادة، الشخص المعنوي بالإجرام هي السارق والقاتل والحرامي وغير ذلك. من الطريقة التي تستعمل لتحليل البيانات في الأفعال الجريمة يعني طريقة تحليل "Cluster" كانت احدى التقنيات في تحليل متعدد التشكيلة بغرض ليجمع بيانات المرصاد أو المتغيرات إلى الفرق تملك قرب المسافة ودرجة التشابه السابق. أما أهداف التجميع هو معرفة الشرطة المحلية تملك القرب أو التشابه في نوع الجريمة، ثم لتتفحص الجريمة التي حدثت في إندونيسيا كثيرا. في تحليل Cluster هناك ثلاث طرق، هي طريقة هرمية Hierarchy وطريقة غير هرمية NoHierarchy وطريقة المجموعة مع الخطوتين two. step clusters كانت طريقة هرمية Hierarchy خمس طرق وهي ربط كامل Complete Linkage، ربط متوسط (Linkage Average)، ربط واحد (Single Linkage)، Ward، Centroid، Linkage. في هذا البحث استخدام طريقة Average Linkage. أما طريقة ربط متوسط تستخدم متوسط القيمة على المتغيرات التي يتم تطبيقها في تجميع عدد الجريمة في إندونيسيا في عام ٢٠١٩. أما نتائج التجميع في هذا البحث فكانت ثلاث مجموعات هي المجموعة الأولى تتكون من اتشبه، رياو، جامبي، بنجكولو، لامونج، بانغكا بليتوغ، رياو، متروجايا، جاوى الغربية، و يوغياكرتا، و بالي، نوسا تنغارا الغربية، كلمنتان الوسطى، كلمنتان الجنوبية، كلمنتان الشرقية، كلمنتان الشمالية، سولاويسي الجنوبية، سولاويسي الوسطى، سولاويسي جنوب الشرقية، غارنتالوا، سولاويسي الغربية، مالوكو الجنوبية. تتكون المجموعة الثانية من سومطرة الشمالية وسومطرة الغربية وسومطرة الجنوبية وجاوى الوسطى وجاوى الشرقية وبانتن ونوسا تنغارا الشرقية وكلمنتان الغربية ومالوكو وبابوا الغربية وبابوا. وتتكون المجموعة الثالثة من سولاويسي الجنوبية فقط.

## BAB 1

### PENDAHULUAN

#### 1.1 Latar Belakang

Ilmu statistika merupakan ilmu yang mempelajari bagaimana merencanakan, mengumpulkan, menganalisis, menginterpretasi, dan mempresentasikan data. Kumpulan data statistik dapat digunakan untuk menyimpulkan atau mendeskripsikan data. Sebagian besar konsep dasar statistik mengasumsikan teori probabilitas. Ilmu statistika banyak diterapkan dalam berbagai ilmu alam, sosial, pendidikan, dan bisnis. Ilmu statistika juga digunakan dalam pemerintahan untuk berbagai macam tujuan. Ilmu statistika memiliki peranan penting dalam perkembangan ilmu pengetahuan dan teknologi. Ilmu statistika juga memegang peranan penting dalam penelitian, baik dalam penyusunan model, perumusan hipotesis, pengembangan alat, instrumen pengumpulan data, penyusunan desain penelitian, penentuan sampel, dan dalam analisis data (Irianto, 2004).

Berdasarkan berbagai pengertian statistika tersebut, maka dalam penerapannya penelitian ini, ilmu statistika akan diaplikasikan dalam kasus kriminalitas yang terjadi di Indonesia tahun 2019. Menurut BPS (2019), publikasi statistik kriminalitas ini menyajikan gambaran umum mengenai tingkat dan perkembangan kriminalitas di Indonesia. Informasi yang disajikan mencakup tiga pendekatan utama statistik kriminal, yakni pendekatan pelaku, korban, dan kewilayahan. Sebagaimana yang sudah tertulis pada undang-undang mengenai kriminalitas yakni pada pasal 333 KUHP yang berisi tentang perampasan hak kemerdekaan orang akan dikenai pidana sesuai jenis pelanggarannya.

Pada ilmu statistika, apabila variabel yang digunakan lebih dari dua, maka akan semakin rumit untuk melakukan analisis statistik. Salah satu metode yang dapat mengatasi hal tersebut adalah analisis multivariat. Analisis multivariat merupakan metode statistik yang memungkinkan melakukan penelitian terhadap lebih dari dua variabel. Salah satu teknik analisis multivariat adalah analisis

*cluster*. Analisis *cluster* merupakan cabang ilmu statistik yang digunakan untuk mengelompokkan obyek-obyek tertentu berdasarkan kesamaan karakter menjadi beberapa kelas. Proses pengelompokan tersebut dapat menggunakan dua metode yaitu *cluster* hirarki dan *cluster non* hirarki. Analisis *Cluster hirarki* (*Hierarchical clustering*) adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih obyek yang memiliki kesamaan paling dekat, dari yang paling mirip sampai yang paling tidak mirip. Berbeda dengan *cluster* hirarki, metode *Cluster non* hirarki (*Non-hierarchical clustering*) justru dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan, kemudian proses *cluster* dilakukan tanpa mengikuti proses hirarki. Metode ini biasa disebut dengan *K-Means Clustering* (Anggara, 2016). Analisis *cluster* terbagi menjadi tiga yaitu *cluster* hirarki, *non* hirarki, dan *two step cluster*. *Two step cluster* digunakan untuk mengklasterkan data dalam jumlah yang sangat besar (Bacher, 2004).

Pada Al-Qur`an juga telah dijelaskan mengenai pengelompokan yang dituangkan dalam surat Al-Waqi`ah ayat 7-13, yang artinya “*Dan kamu menjadi tiga golongan [7], yaitu golongan kanan, alangkah mulianya golongan kanan itu [8], dan golongan kiri, alangkah sengsaranya golongan kiri itu [9], dan orang-orang yang paling dahulu (beriman), merekalah yang paling dahulu (masuk surga) [10], mereka itulah orang yang dekat (kepada Allah) [11], berada dalam surga kenikmatan [12], segolongan besar dari orang-orang yang terdahulu [13]*”(QS. Al-Waqi`ah/56:7-13).

Makna dari Surat Al-Waqi`ah ayat 7-13 menurut Ibnu Katsir adalah manusia pada hari kiamat dibagi menjadi tiga golongan. Satu golongan berada di sebelah kanan `Arsy, golongan ini diberi kitab catatan amal dari tangan kanan kemudian mereka digiring ke dalam golongan kanan. Mereka adalah mayoritas penduduk syurga. Golongan kedua berada di sebelah kiri `Arsy, golongan ini diberi kitab catatan amal dari tangan kiri kemudian mereka digiring ke dalam golongan kiri. Mereka adalah mayoritas penduduk neraka. Golongan ketiga yaitu golongan yang dekat dihadapan Allah. Mereka inilah golongan yang lebih dekat, lebih beruntung, lebih khusus daripada golongan kanan, karena mereka adalah

pemimpin dari golongan kanan. Di tengah-tengah mereka terdapat para rasul, para nabi, orang-orang shiddiq dan orang-orang yang mati syahid. Jumlah golongan ketiga ini jauh lebih kecil dari golongan kanan.

Penelitian sebelumnya membahas analisis *cluster* adalah penelitian berjudul “Analisis *cluster* hirarki untuk pengelompokan dan pemetaan kabupaten/kota di Provinsi Jawa Timur berdasarkan indikator IPM”, terdapat 5 kelompok dengan metode *complete Linkage*, *single linkage*, dan *average linkage* dari 38 kabupaten/kota. Maka dari itu, pada penelitian ini menggunakan metode analisis *cluster* hirarki untuk kasus kriminalitas di Indonesia tahun 2019 sehingga akan dihasilkan kelompok-kelompok berbeda yang menunjukkan karakteristik masing-masing kelompok tersebut. Analisis *cluster* dilakukan setelah dilakukan analisis kasus kriminalitas di Indonesia tahun 2019. Analisis *cluster* ini akan menghasilkan kelompok kejahatan apa yang paling dominan terjadi di Indonesia sehingga sebagai warga negara Indonesia bisa mengantisipasi dengan pencegahan tindakan kejahatan tersebut agar semakin berkurang.

Berdasarkan uraian di atas maka penulis mengangkat permasalahan serta menyusunnya dalam sebuah penelitian yang berjudul “Analisis *Cluster* Hirarki Metode *Average Linkage* Berdasarkan Kriminalitas di Indonesia Tahun 2019”. Penelitian ini diharapkan dapat menambah pengetahuan sekaligus informasi kepada masyarakat umum mengenai kondisi kasus kriminalitas di Indonesia tahun 2019, sehingga tindakan kejahatan tersebut dapat dicegah atau diminimalisir dari sebelumnya.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini, yaitu:

1. Bagaimana analisis *cluster* jika digunakan untuk mengidentifikasi jumlah kriminalitas di Indonesia tahun 2019?

2. Bagaimana aplikasi analisis *cluster* dalam mengidentifikasi jumlah kriminalitas di Indonesia tahun 2019?
3. Apa saja variabel-variabel yang mencirikan tiap kelompok berdasarkan jumlah kriminalitas di Indonesia tahun 2019?

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan yang ingin dicapai adalah penelitian ini, yaitu:

1. Mengetahui analisis *cluster* jika digunakan untuk menganalisis jumlah kriminalitas di Indonesia tahun 2019.
2. Mengetahui aplikasi analisis *cluster* dalam mengidentifikasi jumlah kriminalitas di Indonesia tahun 2019 dengan metode *average linkage*.
3. Mengetahui variabel-variabel yang mencirikan jumlah kriminalitas di Indonesia tahun 2019.

### 1.4 Manfaat Penelitian

Adapun manfaat yang dapat diberikan dalam penelitian ini adalah:

1. Untuk menambah wawasan mengenai analisis *cluster* jika digunakan dalam data jumlah kriminalitas di Indonesia tahun 2019.
2. Untuk menambah wawasan aplikasi analisis *cluster* dalam mengidentifikasi jumlah kriminalitas di Indonesia tahun 2019 dengan metode *average linkage*.
3. Untuk menambah wawasan mengenai variabel-variabel yang mencirikan jumlah kriminalitas di Indonesia Tahun 2019.

## 1.5 Batasan Masalah

Adapun batasan masalah yang ada dalam penelitian:

1. Metode yang dipakai dalam pembentukan *cluster* adalah metode *Average Linkage*.
2. Dalam melakukan analisis peneliti tidak melakukan standarisasi data karena data yang tersedia memiliki ragam dan satuan yang sama.

## 1.6 Sistematika Penulisan

Adapun langkah-langkah yang akan dilakukan dalam penelitian ini berdasarkan suatu sistematika secara garis besar dibagi menjadi lima bab, yaitu:

### Bab I Pendahuluan

Meliputi latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

### Bab II Kajian Pustaka

Berisi mengenai teori-teori yang berhubungan dengan pembahasan antara lain analisis *cluster*, tingkat pendidikan, dan pandangan Islam terhadap *cluster*.

### Bab III Metode Penelitian

Berisi pendekatan penelitian, sumber data, variabel penelitian, dan analisis data.

### Bab IV Pembahasan

Pada bab ini berisi tentang pembahasan pengelompokan jumlah kejahatan menurut jenis kejahatan dan kepolisian daerah di Indonesia tahun 2019 dengan menggunakan analisis *cluster*.

Bab V Penutup

Berisi kesimpulan dan saran.



## BAB II

### KAJIAN PUSTAKA

#### 2.1 Analisis Cluster

Analisis *cluster* merupakan salah satu teknik dalam analisis multivariat yang mempunyai tujuan utama untuk mengelompokkan obyek-obyek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya (Nugroho, 2008). Analisis *cluster* merupakan suatu kelas teknik yang dipergunakan untuk mengklasifikasi obyek atau kasus (responden) ke dalam kelompok yang relatif homogen, yang disebut klaster (*cluster*). Obyek atau kasus dalam setiap kelompok cenderung mirip satu sama lain dan berbeda jauh (tidak sama) dengan obyek dari klaster lainnya. Analisis klaster juga disebut analisis klasifikasi atau taksonomi numerik (*numerical taxonomy*) (Supranto, 2004).

Analisis *cluster* mengklasifikasi obyek sehingga setiap obyek yang paling dekat kesamaannya dengan obyek lain berad dalam *cluster* yang sama. Secara logika, *cluster-cluster* yang baik adalah *cluster* yang terbentuk memiliki (Singgih Santoso, 2002):

- a. Homogenitas (kesamaan) internal yang tinggi antar anggota dalam satu *cluster* (*Within-Cluster*).
- b. Heterogenitas (perbedaan) eksternal yang tinggi antar anggota dalam satu *cluster* (*Between-cluster*).

Fokus dari analisis *cluster* adalah membandingkan obyek berdasarkan variabelnya. Kebanyakan *cluster* tidak didukung dengan suatu penalaran ekstenfif, melainkan berdasarkan pada algoritma (*algorithm*).

## 2.2 Prosedur Analisis *Cluster*

Analisis *cluster* terdiri dari beberapa proses dasar, yaitu (Supranto, 2004):

### a. Merumuskan Masalah

Merumuskan masalah analisis *cluster* merupakan hal yang sangat penting karena pemilihan variabel-variabel yang akan dipergunakan untuk pengklasteran (pembentuk kluster). Memasukkan satu atau dua variabel yang tidak relevan dengan masalah pengklasteran akan mendistorsi hasil pengklasteran yang kemungkinan besar sangat bermanfaat.

### b. Memilih Ukuran Jarak atau Similaritas

Tujuan pengklasteran ialah untuk mengelompokkan obyek yang mirip dalam kluster yang sama, maka beberapa ukuran diperlukan untuk mengakses seberapa mirip atau berbeda obyek-obyek tersebut. Pendekatan yang paling biasa ialah mengukur kemiripan dinyatakan dalam jarak (*distance*) antara pasangan obyek. Jika jarak antara obyek satu dengan yang lain lebih pendek maka akan memiliki kemiripan yang kuat dibandingkan dengan pasangan dengan jarak yang lebih panjang

### c. Standarisasi Data

Proses standarisasi dilakukan apabila diantara variabel-variabel yang diteliti terdapat perbedaan ukuran yang besar. Perbedaan satuan yang mencolok dapat mengakibatkan perhitungan pada analisis *cluster* menjadi tidak valid. Perlu dilakukan proses standarisasi dengan melakukan transformasi (standarisasi pada data asli sebelum dianalisis lebih lanjut). Transformasi dilakukan terhadap variabel yang relevan ke dalam bentuk *Z-score*.

### d. Memilih Suatu Prosedur Pengklasteran

Data yang dianggap mempunyai satuan yang sangat berbeda diseragamkan, dan metode kluster ditentukan (misalnya dipilih *Average Linkage*), kemudian dilakukan pengelompokan data.

e. Melakukan Interpretasi Terhadap *Cluster* Yang Telah Terbentuk

Tahap interpretasi meliputi pengujian tiap *cluster* dalam terminology untuk menamai dan menandai dengan suatu label yang secara akurat dapat menjelaskan *cluster*. Proses ini dimulai dengan suatu ukuran yang sering digunakan yaitu *Centroid Cluster*. Membuat profil dan interpretasi *cluster* tidak hanya untuk memperoleh suatu gambaran saja, melainkan untuk menyediakan rata-rata untuk menilai korespondensi pada *cluster* yang terbentuk serta profil *cluster* memberikan arahan bagi signifikansi praktis.

f. Melakukan Validasi dan *Profiling Cluster*

Proses validasi bertujuan untuk menjamin bahwa solusi yang dihasilkan dari analisis *cluster* dapat mewakili populasi dan dapat digeneralisasi untuk obyek lain. Pendekatan ini membandingkan solusi *cluster* dan melalui hasil korespondensi. *Cluster* yang terbentuk jika sudah valid, maka akan dilakukan proses *profiling* untuk menjelaskan karakteristik setiap *cluster* berdasarkan profil tertentu. Titik beratnya pada katakteristik yang secara signifikan berbeda antar *cluster* dan memprediksi anggota dalam suatu *cluster analysis* berakhir setelah dilalui. Hasil *cluster analysis* dapat digunakan berbagai kepentingan sesuai dengan materi yang dianalisis.

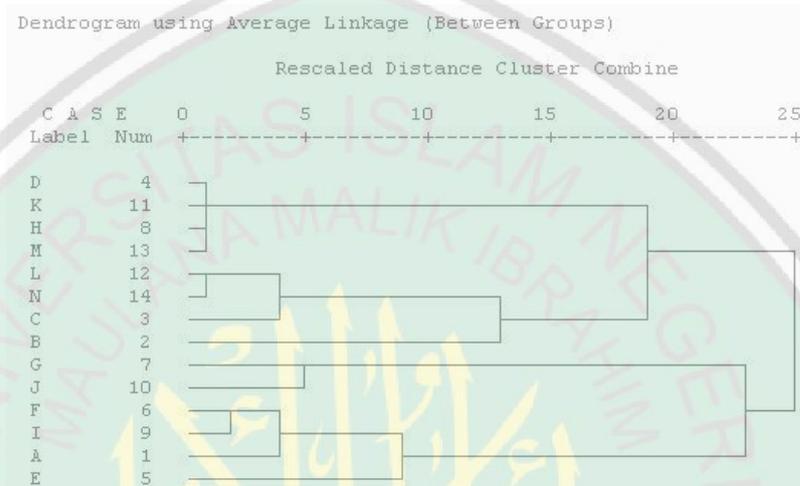
## 2.3 Metode Pengelompokan

Analisis *cluster* terdapat banyak metode untuk pengelompokan observasi ke dalam *cluster*. Namun, secara umum metode *cluster* dapat dikelompokkan menjadi dua yaitu metode hirarki (*Hierarchical Clustering Method*) dan analisis non hirarki (*Nonhierarchical Clustering Method*).

### 2.3.1 Metode Hirarki (Hierarchical Method)

Metode hirarki (*hierarchical method*) adalah suatu metode pada analisis *cluster* yang membentuk tingkatan tertentu seperti pada struktur pohon karena

proses pengklasteran dilakukan secara bertahap atau bertingkat. Hasil Pengklasteran dengan metode hirarki dapat disajikan dalam bentuk dendrogram. Dendrogram adalah representasi visual dari langkah-langkah dalam analisis *cluster* yang menunjukkan bagaimana *cluster* terbentuk dan nilai koefisien jarak pada setiap langkah (Simamora, 2005).



Gambar 2.1 contoh dendrogram *average linkage*

Kelebihan metode *cluster* hirarki, sebagai berikut:

1. Mempercepat pengolahan data.
2. Menghemat waktu karena data yang diinputkan membentuk hirarki (tingkatan).

Kelemahan metode *cluster* hirarki, sebagai berikut:

1. Perbedaan ukuran jarak yang digunakan
2. Terdapat variabel yang tidak relevan

Tahap-tahap pengklasteran data dengan menggunakan metode hirarki adalah (Gudono, 2011):

- 1) Tentukan  $k$  sebagai jumlah *cluster* yang ingin dibentuk
- 2) Setiap data obyek dianggap sebagai *cluster* sehingga  $n = N$

- 3) Menghitung jarak antar *cluster*
- 4) Mencari dua *cluster* yang mempunyai jarak antar *cluster* paling minimal dan menggabungkannya (berarti  $N = n - 1$ )
- 5) Jika  $n > k$ , maka kembali ke langkah 3.

Metode-metode yang bisa digunakan dalam metode hierarki adalah metode agglomerative (*agglomerative method*) dan metode divisive (*divisive method*).

a). Metode *Agglomerative*

Metode agglomerative dimulai dengan menganggap setiap obyek adalah sebuah *cluster*. Dua obyek dengan jarak terdekat digabungkan menjadi satu *cluster*. Proses ini berlanjut sampai terbentuk satu *cluster* yang terdiri dari keseluruhan obyek. *Agglomerative* merupakan prosedur pengelompokan *Hirarki* dimana setiap obyek berawal dari *cluster* yang terpisah. *Cluster-cluster* dibentuk dengan mengelompokkan obyek ke dalam *cluster* yang semakin banyak obyek yang menjadi anggotanya. Proses ini dilanjutkan sampai semua obyek menjadi anggota dari *cluster* tunggal (Supranto, 2004).

Metode *agglomerative* ada beberapa macam, yaitu:

1) Metode *Single Linkage*

Menentukan jarak antar *cluster* dengan menggunakan metode *single linkage* dapat dilakukan dengan melihat jarak antar dua *cluster* yang ada, kemudian memilih jarak paling dekat atau aturan tetangga dekat (*nearest neighbor rule*). Langkah-langkah metode *single linkage* (Johnson & Wichem, 1992):

- (a) Menemukan jarak minimum dalam  $D = \{d_{ij}\}$
- (b) Menghitung jarak antara *cluster* yang telah dibentuk pada langkah 1 dengan obyek lainnya.

- (c) Algoritma diatas dengan jarak-jarak antara (*IJ*) dan *cluster K* yang lain dihitung dengan cara:

$$d_{(ij)k} = \min(d_{IK}, d_{JK}) \quad (2.1)$$

Hal ini besaran-besaran  $d_{IK}$  dan  $d_{JK}$  masing-masing adalah jarak terpendek antara *cluster-cluster I* dan *K* dan juga *cluster-cluster J* dan *K*. Hasil dari pengclusteran menggunakan metode single linkage dapat ditampilkan secara grafis dalam bentuk dendogram (diagram pohon).

#### 2) Metode *Complete Linkage* (*farthest-neighbour method*)

Metode *complete linkage*, jarak antar *cluster* ditentukan oleh jarak terjauh (*farthest-neighbour*) antara dua obyek dalam *cluster* yang berbeda.

$$d_{(IJ)K} = \max(d_{IJ}, d_{JK}) \quad (2.2)$$

Dimana  $d_{IJ}$  dan  $d_{JK}$  masing-masing adalah jarak antara anggota yang paling jauh dari *cluster I* dan *J* serta *cluster J* dan *K* (Johnson & Wichem, 1992).

#### 3) Metode *Centroid*

*Centroid* adalah rata-rata semua obyek dalam *cluster*. Pada metode ini, jarak antar*cluster* adalah jarak antar *centroid*. Centroid baru dihitung ketika setiap kali obyek digabungkan, sehingga setiap kali anggtanya bertambah maka centroidnya akan berubah. Jarak antar *cluster* pada metode ini adalah jarak antar centroid (Johnson & Wichem, 1992).

#### 4) Metode *Average Linkage*

Pada metode *average linkage*, jarak antara dua *cluster* dianggap sebagai jarak rata-rata antara semua anggota dalam satu *cluster* dengan semua anggota *cluster* lain.

$$d_{(IJ)K} = \frac{\sum a \sum b \sum ab}{N_{IJ}N_K} \quad (2.3)$$

Dimana:

$d_{ab}$ : Jarak antara obyek  $i$  pada *cluster* ( $IJ$ ) dan obyek  $b$  pada *cluster*  $K$

$N_{IJ}$  : Jumlah item pada *cluster* ( $IJ$ ).

$N_K$  : Jumlah item pada *cluster* ( $IJ$ ) dan  $K$ .

#### 5) Metode *Ward*

Metode varians bertujuan untuk memperoleh *cluster* yang memiliki varians internal *cluster* yang sekecil mungkin. Metode varians yang umum dipakai adalah metode *ward* dimana rata-rata untuk setiap *cluster* dihitung. Kemudian, dihitung jarak *Euclidean* antara setiap obyek dan nilai rata-rata itu, lalu jarak itu dihitung semua. Pada setiap tahap, dua *cluster* yang memiliki kenaikan ‘sum of squares  $I$  dalam *cluster*’ yang terkecil digabungkan (Simamora, 2005). Metode *ward* adalah metode *cluster* hirarki yang bersifat *agglomerative* untuk memperoleh *cluster* yang memiliki varian internal sekecil mungkin.

#### b). Metode *Devisive*

Metode *Devisive* termasuk dalam analisis *cluster* hirarki. Pada setiap langkahnya, metode *devisive* terjadi penambahan kelompok kedalam dua nilai terkecil, sampai akhirnya semua elemen terklasterkan. Hal ini menjelaskan bahwa *cluster* hirarki dibangun dalam  $n - 1$  langkah ketika data mengandung  $n$  objek. Teknik *devisive* merupakan proses pengklasteran yang didasarkan pada persamaan nilai rata-rata antar objek. Jika sebuah objek memiliki persamaan nilai rata-rata terbesar maka objek tersebut akan terpisah dan berubah menjadi *splinter group*.

Perhitungan teknik *divisive* dilihat dari perbedaan atau selisih antara persamaan nilai rata-rata dengan nilai elemen matrik yang telah menjadi *splinter group*. Jika selisih nilai antara persamaan nilai rata-rata dengan nilai elemen matrik *solinter group* bernilai *negative*, maka perhitungan terhenti sehingga harus

dibuat matrik baru untuk mendapatkan klaster lain. Perhitungan ini terus dilakukan sampai semua objek terpisah.

### 2.3.2 Metode Non-Hirarki (Non Hierarchical Method)

Analisis klaster dengan Metode *non* hirarki merupakan metode *cluster* yang menentukan jumlah *cluster* secara manual (Baroroh, 2012). Analisis *non* hirarki sering disebut sebagai metode *k-means*. Prosedure pada metode *non* hirarki dimulai dengan memilih sejumlah nilai *cluster* awal sesuai dengan jumlah yang diinginkan, kemudian obyek pengamatan digabungkan ke dalam *cluster-cluster* tersebut. Metode *non* hirarki ini meliputi metode *sequential threshold*, *parallel threshold*, dan *optimizing partitioning* sebagai berikut (Gudono, 2011):

#### 1) Metode *sequential threshold*

Metode *sequential threshold* dimulai dengan pemilihan satu *cluster* dan menempatkan semua obyek yang berada pada jarak terdekat ke dalam *cluster* tersebut. Jika semua obyek yang berada pada ambang batas tertentu telah dimasukkan, kemudian *cluster* yang kedua dipilih dan menempatkan semua obyek yang berada pada jarak terdekat ke dalamnya. Kemudian *cluster* ketiga dipilih dan proses dilanjutkan seperti yang sebelumnya.

#### 2) Metode *parallel Threshold*

Metode *parallel threshold* mempunyai prinsip yang sama dengan metode *sequential threshold*, hanya saja pada metode *parallel threshold* dilakukan pemilihan terhadap beberapa obyek awal *cluster* sekaligus dan kemudian melakukan penggabungan obyek ke dalamnya secara bersamaan. Pada saat proses berlangsung, jarak terdekat dapat ditentukan untuk memasukan beberapa obyek ke dalam *cluster-cluster*.

#### 3) Metode *Optimization*

Metode *Optimization* hamper mirip dengan metode *Sequential Threshold* dan metode *Parallel Threshold* yang membedakan adalah metode *optimization*

ini memungkinkan untuk menempatkan kembali obyek-obyek ke dalam *cluster* yang lebih dekat atau dengan melakukan optimasi pada penempatan obyek yang ditukar untuk *cluster* lainnya dengan pertimbangan kriteria optimasi. Ada dua masalah utama pendekatan *non* hirarki yaitu (Simamora, 2005):

- a. Jumlah atau banyaknya *cluster* harus ditentukan terlebih dahulu.
- b. Pemilihan pusat *cluster* tidak menentu (pasti). Seterusnya, hasil pengklasteran tergantung pada bagaimana pusat *cluster* dipilih. Banyak program yang dimulai dengan memilih kasus pertama  $k$  ( $k$  = jumlah *cluster*) sebagai pusat *cluster* awal. Jadi, hasil pengklasteran tergantung pada observasi data. Dibalik segala kekurangan itu, metode ini dapat dilakukan dengan cepat dan sangat bermanfaat kalau jumlah observasi besar.

### 2.3.3 Metode Two Step Cluster

Metode *two step cluster* adalah metode yang didesain untuk menangani jumlah obyek yang besar, terutama pada masalah obyek yang mempunyai peubah campuran, kontinu dan kategorik. Jarak antara dua gerombol didefinisikan sebagai jarak antar pusat dari masing-masing gerombol tersebut. Pusat dari suatu gerombol adalah vektor dari rata-rata masing-masing peubahnya. Jarak yang digunakan dalam metode *two step cluster* adalah jarak *Log-Likelihood* dan jarak Euclidean. Prosedur penggerombolan dengan metode *two step cluster* mempunyai dua tahapan yaitu tahap *preclustering* (penggerombolan awal) obyek ke dalam *subcluster-subcluster* kecil dan tahap penggerombolan akhir (Bacher, 2004).

## 2.4 Metode Menghitung Jarak

Menurut Anderberg tahun 1973 bahwa ukuran jarak dibutuhkan untuk setiap pasang obyek yang akan dikelompokan. Beberapa metode pengukuran jarak antar dua obyek, yaitu:

### 1) Jarak Euclidean

Jarak ini merupakan jarak yang umum digunakan dan dapat digunakan apabila semua peubahnya berskala kontinu. Jarak ini harus memenuhi asumsi bahwa peubah-peubah yang diamati tidak berkorelasi dan antar peubah memiliki satuan yang sama. Metode ini, pengukuran jarak dilakukan dengan menghitung akar kuadrat dari penjumlahan kuadrat selisih dari nilai masing-masing peubah. Jarak Euclid dapat dirumuskan sebagai berikut:

$$d_{i,j} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.4)$$

dengan:

$d_{i,j}$  = jarak antara obyek  $i$  dengan obyek  $k$

$x_{ik}$  = nilai obyek  $i$  pada peubah ke- $k$

$x_{jk}$  = nilai obyek  $j$  pada peubah ke- $k$

$p$  = banyaknya peubah yang diamati

### 2) Jarak Manhattan (*City Block*/Minkowski)

Jarak ini merupakan bentuk umum dari jarak euclidean. Jarak Manhattan digunakan jika peubah yang diamati berkorelasi atau tidak saling bebas. Dalam metode ini, pengukuran jarak dilakukan dengan menghitung jumlah absolut perbedaan obyek untuk masing-masing peubah. Jarak Manhattan dapat dirumuskan sebagai berikut:

$$d_{i,j} = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (2.5)$$

dengan:

$d_{i,j}$  = jarak antara obyek  $i$  dengan obyek  $k$

$d_{ik}$  = nilai obyek  $i$  pada peubah ke- $k$

$x_{jk}$  = nilai obyek  $j$  pada peubah ke- $k$

$p$  = banyaknya peubah yang diamati

### 3) Jarak Mahalonobis

Jarak ini berguna dalam menghilangkan atau mengatasi perbedaan skala pada masing-masing peubah. Jarak mahalonobis dapat dirumuskan sebagai berikut:

$$d_{i,j} = \sqrt{(x_{ik} - x_{jk})'S^{-1}(x_{ik} - x_{jk})} \quad (2.6)$$

dengan:

$d_{i,j}$  = jarak antara obyek  $i$  dengan obyek  $k$

$d_{ik}$  = nilai obyek  $i$  pada peubah ke- $k$

$x_{jk}$  = nilai obyek  $j$  pada peubah ke- $k$

$S$  = matriks kovarian

### 4) Jarak *Log-Likelihood*

Jarak ini digunakan untuk peubah berskala kontinu dan kategorik. Jarak antara gerombol jarak antara obyek  $i$  dengan obyek  $j$  dengan gerombol jarak antara obyek  $i$  dengan obyek  $s$  dapat dirumuskan sebagai berikut:

$$d(j, s) = \xi_j + \xi_s - \xi_{(j,s)} \quad (2.7)$$

dengan:

$$\xi_j = -N \left( \sum_{k=1}^{K^A} \frac{1}{2} \log (\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) - \sum_{k=1}^{K^B} \sum_{i=1}^{L_k} \frac{N_{jkl}}{N_j} \log \left( \frac{N_{jkl}}{N_j} \right) \right)$$

$$\xi_s = -N \left( \sum_{k=1}^{K^A} \frac{1}{2} \log (\hat{\sigma}_k^2 + \hat{\sigma}_{sk}^2) - \sum_{k=1}^{K^B} \sum_{i=1}^{L_k} \frac{N_{skl}}{N_j} \log \left( \frac{N_{skl}}{N_j} \right) \right)$$

$$\xi_{(js)} = -N \left( \sum_{k=1}^{K^A} \frac{1}{2} \log (\hat{\sigma}_k^2 + \hat{\sigma}_{(js)k}^2) - \sum_{k=1}^{K^B} \sum_{i=1}^{L_k} \frac{N_{(js)kl}}{N_j} \log \left( \frac{N_{(js)kl}}{N_j} \right) \right)$$

$N$  = banyaknya obyek

$N_j$  = jumlah obyek di dalam gerombol  $k$

$N_{jkl}$  = jumlah obyek di gerombol  $j$  untuk peubah kategorik ke- $k$  dengan kategori ke- $l$

$\hat{\sigma}_k^2$  = ragam dugaan untuk peubah kontinu ke- $k$  untuk keseluruhan obyek

$\hat{\sigma}_{jk}^2$  = ragam dugaan untuk peubah kontinu ke- $k$  untuk keseluruhan obyek dalam gerombol  $k$

$K^A$  = banyaknya peubah kontinu

$K^B$  = banyaknya peubah kategorik

$L_k$  = banyaknya kategori untuk peubah kategori ke- $k$

##### 5) Jarak *Chi-Square*

Jarak *chi-square* menghitung antara profil dalam jarak euclidean terboboti menggunakan invers proporsi rata-rata sebagai bobot. Misal  $c_j$  menunjukkan elemen ke- $j$  dari rata-rata profil, yang merupakan kelimpahan proporsi peubah  $j$

dalam seluruh kumpulan data. Maka jarak *chi-square*, dinotasikan  $\chi$ , antara dua obyek dengan profil  $x = [x_1, x_2 \dots x_j]$  dan  $y = [y_1, y_2 \dots y_j]$  didefinisikan sebagai:

$$\chi_{x,y} = \sqrt{\sum_{j=1}^J \frac{1}{c_j} (x_j - y_j)^2} \quad (2.8)$$

## 2.5 Uji Asumsi Analisis Cluster

Proses *clustering* bertujuan untuk mengelompokan objek-objek menjadi kelompok-kelompok yang mempunyai karakteristik yang relatif sama (*homogen*). Selain itu, analisis *cluster* juga digunakan untuk membedakan dengan jelas antara satu kelompok (*cluster*) dengan kelompok lainnya. Adapun manfaat analisis *cluster* adalah untuk menerapkan dasar-dasar pengelompokan secara konsisten, mengembangkan suatu metode pengambilan kesimpulan secara umum dengan berdasarkan fakta-fakta khusus, serta mendeskripsikan karakteristik dari masing-masing kelompok.

Asumsi analisis *cluster* ada tiga yaitu (Hair, 1995):

1. Kecukupan sampel.

Analisis *cluster* perlu dilakukan analisis faktor yang menggambarkan variabel-variabel yang saling berkorelasi dengan kuantitas random dengan menggunakan asumsi yang harus dipenuhi melalui uji *keiser meyer olkin* (KMO) dan Uji Bartlett (Maulidina dan Setiawan, 2012). Nilai KMO diatas 0,5 menunjukkan kecukupan sampel dan nilai Bartlett`s diatas 0,05 menunjukkan ragam yang sama.

Statistik uji:

$$KMO = \frac{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 + \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2} \quad (2.9)$$

Dengan:

$i = 1, 2, 3, \dots, n$  dan  $j = 1, 2, 3, \dots, n$  serta  $i \neq j$

$r_{ij}$  : koefisien korelasi antara variabel  $i$  dan  $j$

$a_{ij}$  : koefisien korelasi parsial antara variabel  $i$  dan  $j$

Tabel 2.1 Karakteristik Nilai KMO

Nilai KMO	Keterangan
0,5 – 1,0	Sangat Bagus
0,7 – 0,8	Bagus
0,6 – 0,7	Cukup Bagus
0,5 – 0,6	Tidak Cukup Bagus
$\leq 0,5$	Tidak Layak

- Ragam sama.
- Tidak ada *multikolinearitas*.

*Multikolinearitas* adalah suatu keadaan dimana terdapat korelasi antar variabel predictor ketika dalam model regresi menggunakan lebih dari satu predictor. Apabila terjadi multikolinearitas pada data akan menyebabkan matriks  $(X^1X)^{-1}$  memiliki determinan sama dengan nol. Hocking tahun 1996 mengemukakan bahwa ada tiga kriteria yang dapat digunakan untuk mendeteksi multikolinearitas. Ketiga kriteria tersebut adalah:

- VIF (*Variance Inflation factors*)

Jika nilai VIF lebih besar dari 10 menunjukkan adanya multikolinearitas antara variabel-variabel predictor. VIF dirumuskan

$$VIF = \frac{1}{1-R_j^2} \quad (2.10)$$

Dengan  $R_j$  adalah keefisien determinasi.

- Koefisien korelasi pearson ( $r_q$ )

Koefisien korelasi pearson merupakan salah satu asumsi analisis *cluster* yang harus dipenuhi yaitu tidak terdapat hubungan antara variabel satu dengan yang lainnya atau tidak terdapat korelasi antar variabel. Untuk mengetahui asumsi tersebut terpenuhi maka dilakukan uji korelasi. Korelasi adalah istilah statistika yang digunakan untuk mengukur hubungan antar variabel yang digambarkan oleh besarnya koefisien korelasi. Kefisien korelasi didefinisikan sebagai ukuran

hubungan linier antar variabel acak  $X_1$  dan  $X_2$  dan dilambangkan dengan  $r$  (Bluman, 2004).

$$r = \frac{n \sum_{i=1}^n x_{i1} x_{i2} - (\sum_{i=1}^n x_{i1})(\sum_{i=1}^n x_{i2})}{\sqrt{[n \sum_{i=1}^n x_{i1}^2 - (\sum_{i=1}^n x_{i1})^2][n \sum_{i=1}^n x_{i2}^2 - (\sum_{i=1}^n x_{i2})^2]}} \quad (2.11)$$

Dimana:

$r$  = koefisien korelasi *pearson* antara dua variabel  $X_1$  dan  $X_2$

$n$  = banyaknya pengamatan

$X_{i1}$  = nilai pengamatan ke- $i$  pada variabel  $X_1$

$X_{i2}$  = nilai pengamatan ke- $i$  pada variabel  $X_2$

Pengujian signifikansi koefisien korelasi *pearson* ( $r$ ) digunakan hipotesis sebagai berikut:

$$H_0: P = 0 \text{ lawan } H_1: P \neq 0$$

Pengujian signifikansi tersebut dibandingkan dengan  $r$ -tabel. Dengan ketentuan apabila  $r_{hitung} \geq r_{tabel(a;n-2)}$  maka  $H_0$  ditolak dan apabila  $r_{hitung} < r_{tabel(a;n-2)}$  maka  $H_0$  diterima.

### 3. Nilai eigen

Multikolinearitas terjadi apabila nilai eigen pada matriks korelasi antar semua variabel prediktor  $< 0,05$ .

Solusi untuk mengatasi adanya multikolinearitas adalah dengan mengeluarkan variabel yang tidak signifikan. Ada atau tidaknya multikolinearitas antar variabel sangat diperhatikan dalam analisis *cluster* karena hal itu berpengaruh, sehingga variabel-variabel yang bersifat multikolinearitas secara eksplisit dipertimbangkan dengan lebih seksama. Hal ini juga untuk menentukan ukuran kemiripan (*similarity*) yang akan digunakan dalam analisis *cluster*. Jika asumsi terpenuhi maka jarak *manhattan* dapat digunakan untuk menentukan ukuran kemiripan.

## 2.6 Proses Dasar Analisis Cluster

Proses dasar analisis *cluster* ini merupakan salah satu proses mendasar yang harus diketahui ketika melakukan pengklasteran. Proses dasar ini bisa meliputi

asumsi data yang bisa diklasterkan, mengukur kesamaan antar objek, maupun memilih metode perhitungan jarak.

### 2.6.1 Mengukur Kesamaan Antar Objek (*Similarity*)

Sesuai prinsip analisis *cluster* yang mengelompokkan objek yang mempunyai kemiripan, proses pertama adalah mengukur seberapa jauh ada kesamaan antar objek. Metode yang digunakan:

1. Mengukur korelasi antar sepasang objek pada beberapa variabel.
2. Mengukur jarak (*Distance*) antara dua objek. Pengukuran ada bermacam-macam, yang digunakan pada penelitian ini adalah metode *manhattan distance*.

### 2.6.2 Average Linkage

Menurut Rachmatin (2014), metode *average linkage* adalah metode pengklasteran yang didasarkan pada jarak rata-rata antar objek. Menurut Supranto (2010), metode *average linkage* merupakan gabungan dari *single linkage* dan *complete linkage*. Pada metode *average linkage* proses pengklasteran dimulai dengan menentukan jarak terdekat dalam matriks *proximity* dan proses penggabungan objek-objek dilakukan berdasarkan nilai rata-rata jarak dari objek-objek tersebut, misalkan  $U$  dan  $V$ , sehingga diperoleh klaster  $(UV)$ . Selanjutnya jarak antara  $(UV)$  dan klaster  $W$  lainnya dapat diperoleh dengan cara sebagai berikut:

$$d_{(UV),W} = \frac{\left[ \sum_{i=1}^n \sum_{k=1}^k d_{ik} \right]}{N_{(UV)}N_{(W)}} \quad (2.12)$$

dimana:

$d_{ik}$  : jarak antara objek  $i$  dalam kelompok  $(UV)$  dan objek  $k$  dalam kelompok  $W$

$N_{(UV)}$  : banyaknya objek-objek dalam kelompok  $(UV)$

$N_{(W)}$  : banyaknya objek-objek dalam kelompok  $W$

## 2.7 Kajian Jenis Kriminalitas di Indonesia

Kriminalitas atau tindak kejahatan adalah tingkah laku yang melanggar hukum dan melanggar norma-norma sosial, sehingga masyarakat menentangnya. Banyak kasus kejahatan terjadi karena beberapa faktor. Faktor penyebab kejahatan antara lain faktor biologik, sosiologik yang terdiri dari faktor-faktor ekonomi (sistem ekonomi, populasi, perubahan harga pasar, krisis moneter, kurangnya lapangan kerja, dan pengangguran), faktor-faktor mental (agama, bancaan, harian-harian, film), faktor-faktor fisik seperti keadaan iklim dan lain-lain serta faktor-faktor pribadi seperti umur, ras, dan nasionalitas, alkohol, dan perang (Kartono, 1999).

Penelitian ini akan dilihat jenis kejahatan yang paling banyak terjadi di Indonesia dengan menggunakan tiga jenis kejahatan yaitu mempekerjakan anak dibawah umur, penculikan dan korupsi. Ketiga jenis kejahatan tersebut termasuk tindakan kriminal yang sangat berdampak negatif terhadap kehidupan bermasyarakat antara lain menimbulkan rasa tidak aman, kecemasan, ketakutan dan kepanikan (Kartono, 1999). Maka dari itu, pencegahan tindakan kriminalitas perlu diupayakan agar tidak merugikan banyak pihak.

## 2.8 Analisis *cluster* dalam Pandangan Islam

*Cluster* mempunyai arti sama dengan kelompok-kelompok. Didalam Islam banyak dijelaskan mengenai hal-hal yang berkaitan dengan pengelompokan. Yang sudah diajarkan dalam Islam adalah mengenai orang-orang yang beriman. Seperti pada firman Allah swt yang artinya “*Sesungguhnya orang-orang yang beriman ialah mereka yang bila disebut nama Allah gemetarlah hati mereka, dan apabila dibacakan ayat-ayat-Nya bertambahlah iman mereka (karenanya), dan hanya kepada Tuhanlah mereka bertawakal, yaitu orang-orang yang mendirikan shalat dan yang menafkahkan sebagian dari rezeki yang Kami berikan kepada mereka. Itulah orang-orang yang beriman dengan sebenar-benarnya. Mereka akan memperoleh beberapa derajat ketinggian di sisi Tuhannya dan ampunan serta rezeki yang mulia*” (QS. Al Anfal/8:2-4).

Makna Surat Al-Anfal ayat 2-4 menurut Ibnu Katsir adalah orang munafik hatinya tidak sedikitpun mengingat Allah dalam menjalankan kewajiban-kewajibannya. Mereka juga tidak beriman kepada ayat-ayat Allah, tidak bertawakkal, tidak sholat saat sendirian, dan tidak menunaikan zakat dalam harta kekayaan mereka. Maka Allah memberitahukan, bahwa mereka bukanlah orang-orang yang beriman. Kemudian Allah mensifati orang-orang beriman apabila disebut nama Allah, bergetarlah hatinya, mereka melaksanakan kewajiban-kewajiban dan apabila dibacakan ayat-ayat Allah, maka bertambahlah imannya (Abu Al-Fida` Ibnu Katsir, 1999).

Surat Al-Anfal ayat 2-4 menjelaskan tentang golongan orang-orang yang beriman dan bertawakkal kepada Allah swt. Apabila orang-orang memiliki salah satu ciri-ciri sama seperti dalam ayat tersebut maka orang-orang tersebut dikelompokkan ke dalam kelompok orang-orang yang memiliki sifat atau ciri-ciri yang sama seperti. Begitu pula pada *cluster*, jika suatu obyek mempunyai kesamaan karakteristik maka obyek tersebut akan berada pada *cluster* yang sama.

## BAB III

### METODOLOGI PENELITIAN

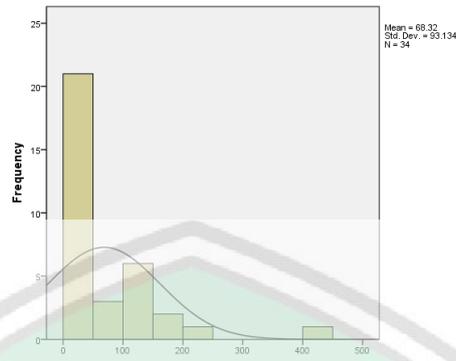
#### 3.1 Pendekatan Penelitian

Pendekatan penelitian yang digunakan pada penelitian ini adalah pendekatan studi literatur deskriptif kuantitatif. Pendekatan dengan studi literatur yaitu dengan mengumpulkan bahan-bahan pustaka yang dibutuhkan oleh peneliti sebagai acuan dalam menyelesaikan penelitian, dan pendekatan deskriptif kuantitatif yaitu dengan menganalisis data yang sudah ada sesuai kebutuhan peneliti. Pendekatan penelitian ini dapat dimulai dengan mendeskripsikan data. Deskripsi data dilakukan terlebih dahulu sebelum melakukan analisis kluster berdasarkan jumlah kriminalitas di Indonesia tahun 2019. Hal tersebut perlu dilakukan untuk melihat gambaran umum dari setiap variabel.

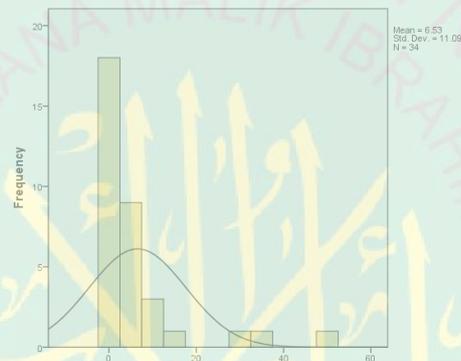
Tabel 3.1 Deskripsi Data

Jenis kejahatan	N	Minimum	Maximum	Sum	Mean	Variance
Mempekerjakan anak dibawah umur	34	0	434	2323	68.32	8673.922
Penculikan	34	0	51	222	6.53	122.984
Korupsi	34	0	38	496	14.59	129.704
Valid N (listwise)	34					

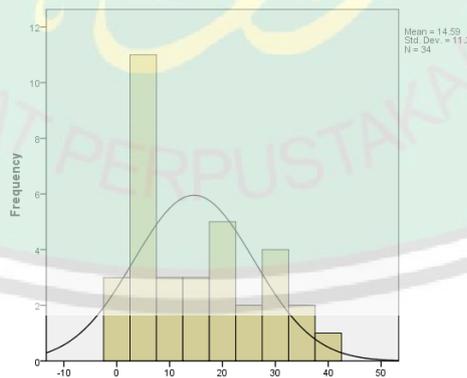
Tabel 3.1 dapat diketahui secara global bahwa rata-rata jumlah kriminalitas di Indonesia tahun 2019 yang mempekerjakan anak dibawah umur sebesar 68.32 kasus dengan jumlah sebesar 2323 kasus dan *variance* sebesar 8673.922 kasus dengan rentang antara 0 sampai 434 kasus. Pada penculikan memiliki rata-rata sebesar 6.53 kasus dengan jumlah sebesar 222 kasus dan *variance* sebesar 122.984 kasus dengan rentang antara 0 sampai 51 kasus. Pada Korupsi memiliki rata-rata sebesar 14.59 kasus dengan jumlah sebesar 496 kasus dan *variance* sebesar 129.704 kasus dengan rentang antara 0 sampai 38 kasus.



Gambar 3.1  
Histogram Mempekerjakan Anak Dibawah Umur



Gambar 3.2  
Histogram Penculikan



Gambar 3.3  
Histogram Korupsi

### 3.2 Sumber Data

Pada penelitian ini data yang digunakan mengenai jumlah kriminalitas di Indonesia tahun 2019, yang bersumber dari Badan Pusat Statistik (BPS) Indonesia. Unit observasi penelitian ini adalah 34 kepolisian daerah di Indonesia.

Tabel 3.2 Data

No	Kepolisian Daerah	Mempekerjakan anak dibawah umur	Penculikan	Korupsi
1	Aceh	12	29	15
2	Sumatera Utara	73	15	29
3	Sumatera Barat	131	1	12
4	Riau	15	1	21
5	Jambi	49	4	19
6	Sumatera Selatan	129	3	22
7	Bengkulu	4	2	24
8	Lampung	2	5	4
9	Kep. Bangka Belitung	33	0	4
10	Kepulauan Riau	0	2	7
11	Metro Jaya	4	34	9
12	Jawa Barat	21	12	28
13	Jawa Tengah	192	7	29
14	DI Yogyakarta	13	1	3
15	Jawa Timur	149	12	35
16	Banten	136	0	5
17	Bali	26	3	21
18	Nusa Tenggara Barat	52	2	5
19	Nusa Tenggara Timur	72	6	18
20	Kalimantan Barat	128	1	29
21	Kalimantan Tengah	3	0	38
22	Kalimantan Selatan	2	0	23
23	Kalimantan Timur	0	0	35
24	Kalimantan Utara	1	0	1
25	Sulawesi Utara	0	51	5
26	Sulawesi Tengah	15	0	8
27	Sulawesi Selatan	434	6	14
28	Sulawesi Tenggara	11	1	3
29	Gorontalo	32	6	5
30	Sulawesi Barat	10	0	1
31	Maluku	193	12	4
32	Maluku Utara	18	2	0
33	Papua Barat	129	0	6
34	Papua	234	4	14

### 3.3 Variabel Penelitian

Adapun variabel yang digunakan dalam penelitian ini, antara lain:

$X_1$  : Mempekerjakan anak dibawah umur

$X_2$  : Penculikan

$X_3$  : Korupsi

### 3.4 Analisis Data

Analisis data merupakan kegiatan yang dilakukan untuk mengubah hasil data dari penelitian menjadi informasi baru yang dapat digunakan dalam membuat kesimpulan. Secara umum analisis data digunakan agar data lebih mudah dipahami dan dapat digunakan sebagai solusi dari suatu permasalahan.

#### 3.4.1 Pengelompokan Jumlah Kriminalitas di Indonesia

Adapun langkah-langkah analisis data untuk melakukan penelitian tentang pengelompokan jumlah kriminalitas di Indonesia tahun 2019 antara lain:

##### 1. Merumuskan Masalah

Tahap ini merupakan tahapan yang paling utama sebelum melakukan analisis *cluster*. Tahap ini adalah tahap pemilihan variabel-variabel yang akan dipergunakan untuk pengklasteran (pembentukan *cluster*). Memasukan satu atau dua variabel yang tidak relevan dengan masalah pengklasteran akan menyebabkan penyimpangan hasil pengklasteran yang kemungkinan besar sangat bermanfaat (Supranto, 2004).

##### 2. Memilih Ukuran Jarak

Menurut Nugroho tahun 2008 bahwa pengukuran jarak untuk mengukur jarak (*Distance*) antara dua obyek terdiri dari jarak *euclid*, kuadrat *euclid*, *city-block* atau *manhattan*, dan *chebychev*. Pada penelitian ini jarak yang akan digunakan adalah *city-block* atau *Manhattan*.

### 3. Asumsi-asumsi

Terdapat dua asumsi yang harus dipenuhi dalam analisis *cluster* yaitu sampel mewakili populasi dan tidak terjadi multikolinearitas. Cara mengetahui apakah sampel yang diambil benar-benar dapat mewakili populasi yang ada dibutuhkan nilai *Kaiser-Meyer Olkin (KMO)*. Nilai KMO kurang dari 0,5 menandakan bahwa sampel yang diambil tidak dapat mewakili populasi yang ada (Machfudhoh, 2013). Sedangkan untuk mendeteksi ada tidaknya multikolinearitas dapat dilakukan dengan menghitung koefisien korelasi sederhana antar variabel bebas, jika nilai mencapai atau melebihi 0,8 maka terjadi multikolinearitas pada data tersebut. Cara mengatasi masalah multikolinearitas pada data adalah dengan *Metode Principal Component Analysis (PCA)* atau analisis komponen utama.

### 4. Standarisasi Data

Standarisasi data perlu dilakukan sebelum dianalisis lebih lanjut. Cara standarisasi data yaitu dengan melakukan transformasi pada data asli, apabila variabel yang diteliti memiliki perbedaan ukuran satuan yang besar. Standarisasi dilakukan terhadap variabel yang relevan ke dalam bentuk *z-score*.

### 5. Memilih Metode Pengklasteran

Metode *cluster* terdiri dari tiga macam yaitu metode hierarki, *non hierarki*, dan *two step cluster*. Namun pada penelitian ini akan digunakan metode *Hirarki* yang berupa metode *average linkage*. Menurut Rachmatin tahun 2014 bahwa metode *average linkage* adalah metode pengklasteran yang didasarkan pada jarak rata-rata antar obyek. Sedangkan menurut Supranto tahun 2010 bahwa metode *average linkage* merupakan gabungan dari *single linkage* dan *complete linkage*.

### 6. Interpretasi output *cluster* yang sudah terbentuk

Menurut Supranto tahun 2010 bahwa interpretasi meliputi pengkajian mengenai centroids yaitu rata-rata nilai obyek yang terdapat dalam *cluster* pada

setiap variabel. Nilai centroid memungkinkan untuk menguraikan setiap *cluster* dengan cara memberikan suatu nama atau label.

#### 7. Melakukan validasi dan profiling *cluster*

Validasi bertujuan untuk menjamin bahwa solusi yang dihasilkan dari *cluster* analisis dapat mewakili populasi dan dapat digeneralisasikan ke obyek lain. Sedangkan profiling data bertujuan untuk menguji kevalidan *cluster* yang terbentuk untuk menjelaskan karakteristik setiap *cluster*.

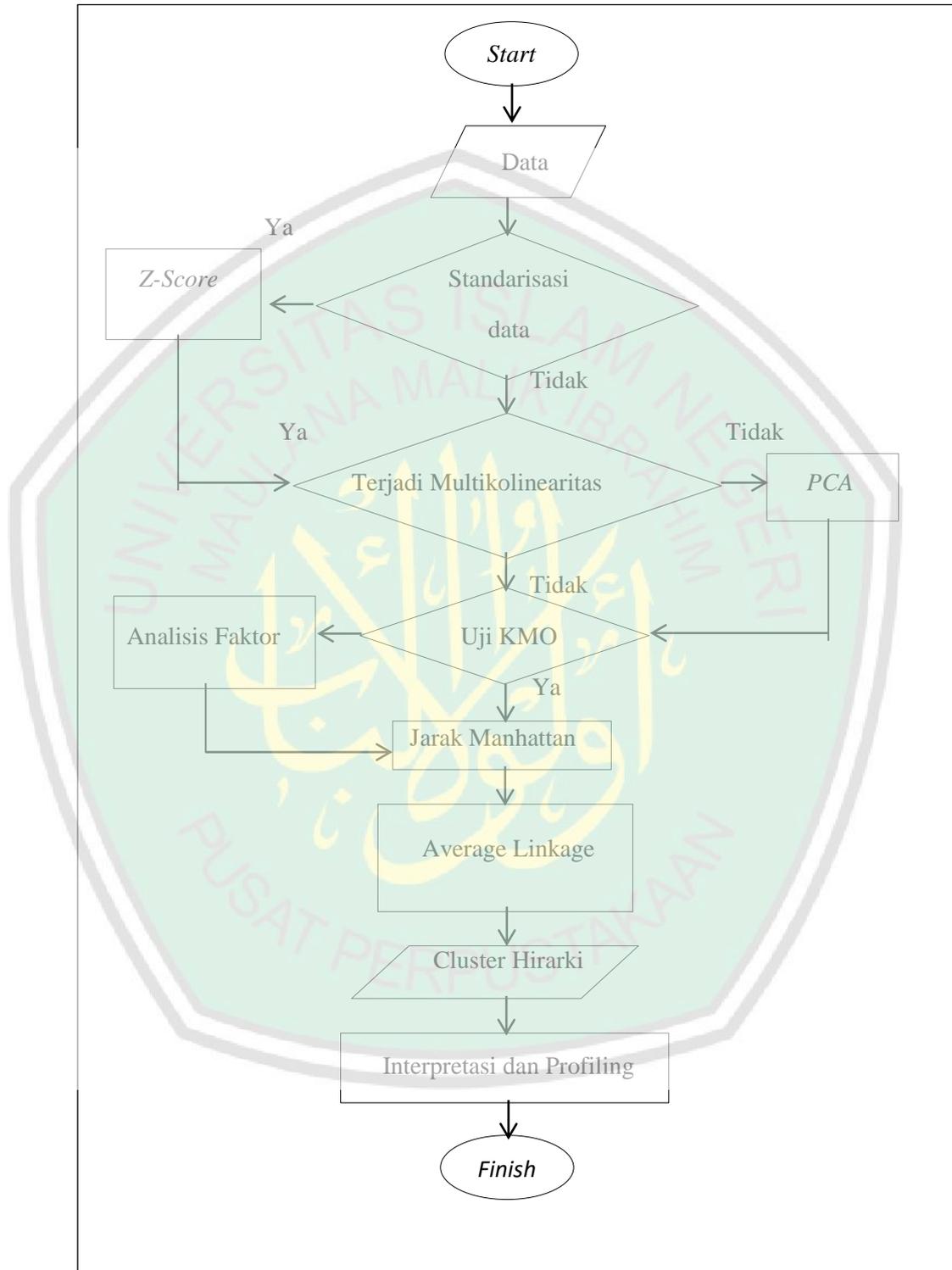
#### 8. Menarik Kesimpulan

##### 3.4.2 Karakteristik Setiap Kelompok

Langkah-langkah dalam menentukan karakteristik setiap kelompok adalah sebagai berikut:

1. Menghitung nilai rata-rata dari setiap variabel yang terdapat dalam masing-masing kelompok yang terbentuk.
2. Menentukan nilai rata-rata yang terbesar dari setiap variabel dalam masing-masing kelompok yang akan menjadi karakteristik kelompok tersebut.
3. Menarik kesimpulan.

### 3.5 Flowchart Analisis Cluster



Tabel 3.2 Keterangan Simbol *Flowchart*

Nama Simbol	Keterangan
<p><i>Terminal point symbol</i></p> 	Simbol yang digunakan sebagai permulaan atau mengakhiri suatu proses.
<p><i>Flow direction symbol</i></p> 	Simbol yang digunakan untuk menghubungkan simbol satu dengan simbol lainnya.
<p><i>Input output symbol</i></p> 	Simbol yang menunjukkan proses input output
<p><i>Decision symbol</i></p> 	Simbol yang digunakan untuk memilih proses atau keputusan berdasarkan kondisi yang ada.
<p><i>Processing symbol</i></p> 	Simbol yang digunakan untuk menunjukkan proses kegiatan yang dilakukan.

## BAB IV

### PEMBAHASAN

#### 4.1 Analisis *cluster* pada Jumlah Kriminalitas

Data kriminalitas sebelum di analisis *cluster* harus diuji dulu apakah data tersebut sudah memenuhi asumsi-asumsi analisis *cluster*. Asumsi analisis *cluster* ada tiga yaitu uji kecukupan sampel, memiliki ragam yang sama, dan datanya *non* multikolinearitas. Jika salah satu dari asumsi tersebut tidak terpenuhi maka data tersebut tidak bisa dianalisis *cluster*. Namun, jika datanya multikolinearitas dapat diatasi dengan menganalisis faktor terlebih dahulu sebelum diklasterkan.

##### 4.1.1 Uji Asumsi Analisis *Cluster*

Uji asumsi analisis *cluster* ada tiga macam yaitu uji kecukupan sampel, ragam sama, dan *non* multikolinearitas. Data yang tidak memenuhi salah satu atau ketiga asumsi tersebut maka data itu tidak dapat di analisis *cluster*. Maka dari itu, pengujian ketiga asumsi tersebut sangatlah penting untuk analisis *cluster*.

##### 1. Uji Kecukupan Sampel dan Ragam

Tabel 4.1 KMO and Bartlett`s Test

<i>Kaiser-Meyer-Olkin Measure of Sampling Adequacy</i>		.523
<i>Bartlett`s Test of Sphericity</i>	<i>Approx. Chi-Square</i>	.547
	<i>Df</i>	3
	<i>Sig.</i>	.098

Tabel 4.1 menjelaskan bahwa nilai KMO pada data yang digunakan sebesar 0.523, artinya nilai tersebut diatas 0.5 yang berarti data tersebut memenuhi kecukupan sampel. Pada Uji *Bartlett`s* nilai *Sig* (*P-Value*) sebesar 0.98, artinya nilai *Sig* tersebut diatas 0,05 (terima  $H_0$ ). Jika nilai *Sig* diatas 0.05 maka data yang digunakan memiliki ragam yang sama.

## 2. Uji Korelasi

Asumsi yang harus di penuhi setelah kecukupan sampel dan ragam adalah tidak adanya hubungan antar variabel (multikolinearitas).

Tabel 4.2 Uji VIF

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	T	P	VIF
<i>Constant</i>	13.944	2.823	4.94	0.000	
Mempekerjakan anak dibawah umur	0.01237	0.02190	0.57	0.576	1.006
Penculikan	-0.0309	0.1839	-0.17	0.868	1.006

Tabel 4.3 Uji VIF

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	T	P	VIF
<i>Constant</i>	7.556	3.425	2.21	0.035	
Mempekerjakan anak dibawah umur	-0.00874	0.02143	-0.41	0.686	1.011
Korupsi	-0.0294	0.1752	-0.17	0.868	1.011

Tabel 4.4 Uji VIF

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	T	P	VIF
<i>Constant</i>	60.29	28.83	2.09	0.045	
Penculikan	-0.611	1.497	-0.41	0.686	1.001
Korupsi	0.824	1.458	0.57	0.576	1.001

Pada tabel 4.2 sampai 4.4 dapat diketahui bahwa nilai VIF sebesar 1.006, 1.011, dan 1.001 artinya nilai tersebut lebih besar dari 0,01 (tidak memiliki multikolinearitas). Berdasarkan tabel 4.1, 4.2, 4.3 dan 4.4 dapat disimpulkan bahwa data yang digunakan memenuhi asumsi analisis *cluster* sehingga data dapat diproses ke tahap selanjutnya. Namun, uji korelasi juga dapat dilakukan dengan cara menghitung koefisien korelasi pearson ( $r$ ) pada persamaan 2.11.

Berikut ini merupakan tahap-tahap uji korelasi:

1. Membuat hipotesis

$$H_0: P = 0 \text{ lawan } H_1: P \neq 0$$

$H_0: P = 0$  ; tidak terdapat korelasi antara variabel  $X_1$  dan  $X_2$

$H_1: P \neq 0$  ; terdapat korelasi antara variabel  $X_1$  dan  $X_2$

2. Memilih taraf signifikan ( $\alpha = 0,05$ )

3. Menghitung nilai koefisien korelasi pearson

$$r = \frac{n \sum_{i=1}^n x_{i1} x_{i2} - (\sum_{i=1}^n x_{i1})(\sum_{i=1}^n x_{i2})}{\sqrt{[n \sum_{i=1}^n x_{i1}^2 - (\sum_{i=1}^n x_{i1})^2][n \sum_{i=1}^n x_{i2}^2 - (\sum_{i=1}^n x_{i2})^2]}}$$

$$r = \frac{27578,73529}{\sqrt{4998,11765} \cdot \sqrt{286239,4412}}$$

$$r = 0,2306$$

Maka nilai koefisien korelasi pearson  $X_1$  dan  $X_2$  adalah 0,2306, sehingga

$$r_{hitung} = 0,2306$$

4. Menentukan titik kritis ( $r_{tabel(a;n-2)} = r_{tabel(0,001;32)} = 0,2869$ )

5. Membuat keputusan apakah  $H_0$  diterima atau ditolak dengan kriteria sebagai berikut:

$H_0$  diterima jika  $r_{hitung} < r_{tabel(a;n-2)}$  dan  $H_0$  ditolak jika  $r_{hitung} \geq r_{tabel(a;n-2)}$ .

6. Membuat kesimpulan

Karena  $r_{hitung} = 0,2306$  lebih kecil dari  $r_{tabel(a;n-2)} = 0,2869$ , maka kesimpulannya adalah  $H_0$  diterima, artinya tidak ada korelasi antara  $X_1$  dan  $X_2$ .

#### 4.2 Penerapan Analisis *cluster* pada Jumlah Kriminalitas

Analisis *cluster* ini baru bisa diaplikasikan pada data setelah ketiga sumsi tersebut terpenuhi. Setelah ketiga asumsi tersebut terpenuhi, maka proses selanjutnya menghitung jarak dari masing-masing variabel. Pada penelitian ini, penulis menggunakan jarak mahattan. Setelah menghitung jarak, maka akan dilakukan pengklasteran. Pengklasteran ini menggunakan nilai rata-rata (*average*). Kemudian, tahap terakhir yaitu validasi dan profiling *cluster* agar hasil *cluster* yang diperoleh akurat.

#### 4.2.1 Jarak Manhattan

Adapun cara menghitung jarak *Manhattan* adalah sebagai berikut:

Perhitungan kemiripan antara kepolisian daerah Aceh dengan Sumatera Utara (data 1 dan 2):

$$d_{i,j} = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

$$d_{1,2} = |x_1 - y_1| + |x_2 - y_2| + |x_3 - y_3|$$

$$d_{1,2} = |12 - 73| + |29 - 15| + |15 - 29|$$

$$d_{1,2} = 89$$

Perhitungan kemiripan antara data 1 dan 2 jarak *manhattan* sebesar 89, kemudian kemiripan antara kepolisian daerah Aceh dengan Sumatera Utara (data 1 dan 3):

$$d_{i,j} = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

$$d_{1,3} = |x_1 - y_1| + |x_2 - y_2| + |x_3 - y_3|$$

$$d_{1,3} = |12 - 131| + |29 - 1| + |15 - 12|$$

$$d_{1,3} = 150$$

Perhitungan kemiripan antara data 1 dan 2 jarak *manhattan* sebesar 150, kemudian kemiripan antara kepolisian daerah Sumatera Utara dengan Sumatera Barat (data 2 dan 3):

$$d_{i,j} = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

$$d_{2,3} = |x_1 - y_1| + |x_2 - y_2| + |x_3 - y_3|$$

$$d_{2,3} = |73 - 131| + |15 - 1| + |29 - 12|$$

$$d_{2,3} = 89$$

Perhitungan kemiripan antara data 2 dan 3 jarak *manhattan* sebesar 89, kemudian kemiripan antara kepolisian daerah Sumatera Utara dengan Riau (data 2 dan 4):

$$d_{i,j} = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

$$d_{2,4} = |x_1 - y_1| + |x_2 - y_2| + |x_3 - y_3|$$

$$d_{2,4} = |73 - 15| + |15 - 1| + |29 - 21|$$

$$d_{2,4} = 80$$

Perhitungan kemiripan antara data 3 dan 4 jarak *manhattan* sebesar 80. Dibawah ini merupakan ringkasan tabel matriks dari jarak manhattan:

Tabel 4.5 Matriks Jarak Manhattan

$D_{man}$	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$	...	$D_{34}$
$D_1$	0.000	89.000	15.000	37.000	66.000	...	248.000
$D_2$	89.000	0.000	89.000	80.000	45.000	...	187.000
$D_3$	150.000	89.000	0.000	125.000	92.000	...	108.000
$D_4$	37.000	80.000	125.000	0.000	39.000	...	229.000
$D_5$	66.000	45.000	92.000	39.000	0.000	...	190.000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$D_{34}$	248.000	187.000	108.000	229.000	190.000	...	0.000

Pada matriks diatas dapat dilihat hasil jarak *manhattan* antara kepolisian daerah Aceh dan Sumatera Utara sebesar 89, sedangkan jarak *manhattan* kepolisian daerah Aceh dan Sumatera Barat sebesar 150, jarak *manhattan* antara kepolisian daerah Aceh dan Riau sebesar 37, jarak *manhattan* kepolisian daerah antara Aceh dan Jambi sebesar 66. Pada jarak tersebut menunjukkan bahwa jarak antara kepolisian daerah Aceh dengan Riau lebih dekat atau mirip daripada jarak kepolisian daerah Aceh dengan Sumatera Utara. Hal ini berlaku untuk semua data, semakin kecil nilai jarak maka semakin tinggi tingkat kemiripan karakteristik antar data.

#### 4.2.2 Metode Average Linkage

Pada penelitian ini menggunakan metode *average linkage*. Metode ini dengan mencari nilai rata-rata antar objek dengan jarak *manhattan* sehingga diperoleh matriksnya. Matrik data kriminalitas dapat dilihat dilampiran. Setelah matriks diketahui, maka langkah selanjutnya mencari data yang paling dekat (mirip) pada matrik jarak *manhattan* kemudian menggabungkannya menjadi sebuah kelompok atau *cluster* baru. Adapun rumus metode *average linkage* sebagai berikut:

$$d_{(UV),W} = \frac{[\sum_{i=1}^n \sum_{k=1}^k d_{ik}]}{N_{(UV)}N_{(W)}}$$

Pengelompokan ini dimlai dari data ke 14 dan 28 karena data tersebut memiliki jarak terdekat sebesar 2,000. Data lengkapnya dapat dilihat di lampiran.

$$\begin{aligned} d_{(14,28)1} &= \text{average}(d_{14,1}; d_{28,1}) \\ &= \text{average}(41.000; 41.000) \\ &= 41.000 \end{aligned}$$

$$\begin{aligned} d_{(14,28)2} &= \text{average}(d_{14,2}; d_{28,2}) \\ &= \text{average}(100.000; 102.000) \\ &= 101.000 \end{aligned}$$

$$\begin{aligned} d_{(14,28)3} &= \text{average}(d_{14,3}; d_{28,3}) \\ &= \text{average}(127.000; 129.000) \\ &= 128.000 \end{aligned}$$

$$\begin{aligned} d_{(14,28)4} &= \text{average}(d_{14,4}; d_{28,4}) \\ &= \text{average}(20.000; 22.000) \\ &= 21.000 \end{aligned}$$

$$\begin{aligned} d_{(14,28)5} &= \text{average}(d_{14,5}; d_{28,5}) \\ &= \text{average}(55.000; 57.000) \\ &= 56.000 \end{aligned}$$

$$d_{(14,28)2} = \text{average}(d_{14,2}; d_{28,2})$$

$$= \text{average}(137.000; 139.000)$$

$$= 138.000$$

$$d_{(14,28)2} = \text{average}(d_{14,2}; d_{28,2})$$

$$= \text{average}(31.000; 29.000)$$

$$= 30.000$$

Perhitungan tersebut dilakukan sampai menghasilkan sebuah matriks baru.

Tabel 4.6 Matriks Jarak Baru

$D_{man}$	1,11,25,4,17,12,7,22,21, 23,8,10,24,14,28,30,26,32,5, 18,9,29	2,19,3,16,33,6,20,15,13,31,34	27
1,11,25,4,17,12,7,22,21, 23,8,10,24,14,28,30,26,32,5, 18,9,29	0.000	163.455	305.818
2,19,3,16,33,6,20,15,13,31,34	163.455	0.000	305.818
27	305.818	305.818	0.000

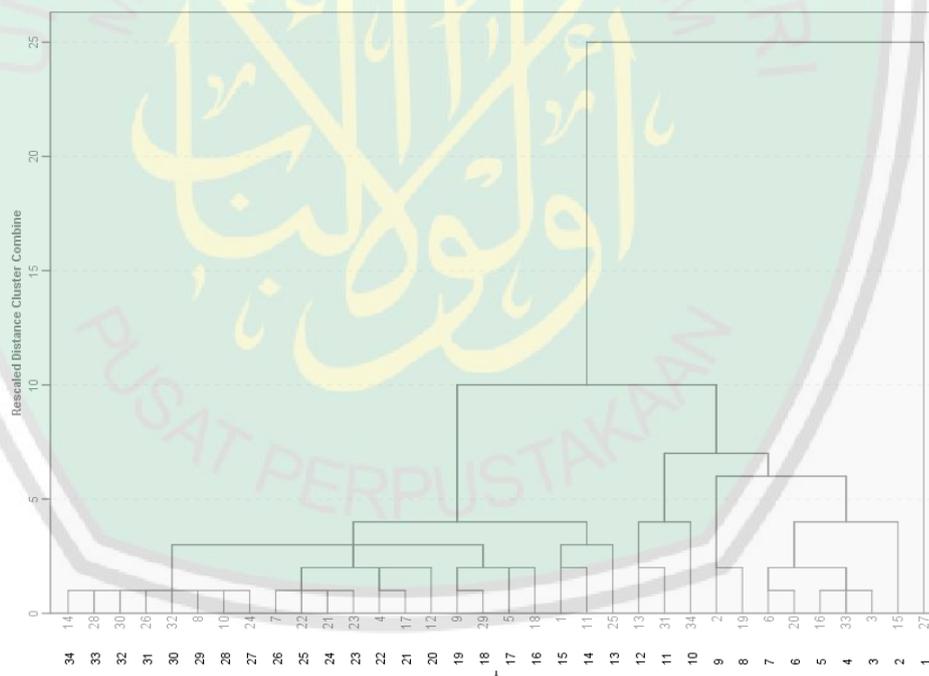
Matriks tersebut merupakan hasil dari perhitungan menggunakan metode penggabungan, yang dimulai dari jarak yang paling mirip atau jarak terdekat. Berikut ini merupakan tabel *distance cluster*.

Tabel 4.7 Cluster Joined

Steps	Similarity Level	Distance Level	Clusters Joined	
1	99.5902	2,000	14	28
2	98.9754	5,000	14	30
3	98.9754	5,000	7	22
4	98.7705	6,000	21	23
5	98.3607	8,000	16	33
6	98.3607	8,000	9	29
7	98.3607	8,000	8	10
8	98.1557	9,000	8	24
9	97.9508	10,000	14	26
10	97.9508	10,000	6	20
11	97.7459	11,000	1	32

12	97.7459	11.000	3	16
13	97.3361	13.000	4	17
14	96.7213	16.000	7	21
15	96.3934	17.600	8	14
⋮	⋮	⋮	⋮	⋮
33	19.3058	393.788	1	27

Pada tabel diatas dapat diketahui nilai *similarity* dari masing-masing kluster sangatlah berdekatan. Masing-masing *cluster* juga memiliki jarak (*Distance*) yang berbeda-beda, jarak ini dihitung dengan metode *manhattan*. Sedangkkn *cluster joined* merupakan kolom yang memiliki *distance* yang sama, pada tabel tersebut bisa dilihat bahwa 14 dengan 28, 14 dengan 30 sampai yang terakhir yaitu 1 dengan 27.



Gambar 4.1 Dendrogram Analisis *Cluster* Metode *Average Linkage*

Pada dendrogram diatas menjelaskan tentang *cluster joined*, dimana hal tersebut didasarkan pada *distance level* yang sama. Penentuan banyaknya *cluster* didasarkan ada jarak terjauh dari *cluster* tersebut. Pengelompokan pada

dendrogram diatas dimulai dengan 2 objek yang memiliki jarak terdekat, dapat dilihat pada langkah-langkah penggelompokan (*Agglomerative*) berikut:

1. Tahap 1 data ke 14 dan 48 bergabung menjadi kelompok 1.
2. Tahap 2 data ke 14 dan 30 bergabung menjadi kelompok 2, sehingga kelompok ini terdiri dari data ke 14, 28, dan 30.
3. Tahap 3 data ke 7 dan 22 bergabung menjadi kelompok 3.
4. Tahap 4 data ke 21 dan 23 bergabung menjadi kelompok 4.
5. Tahap 5 data ke 16 dan 33 bergabung menjadi kelompok 5.
6. Tahap 6 data ke 9 dan 29 bergabung menjadi kelompok 6.
7. Tahap 7 data ke 8 dan 10 bergabung menjadi kelompok 7.
8. Tahap 8 data ke 8 dan 24 bergabung menjadi kelompok 8, sehingga kelompok ini terdiri dari data ke 8, 10, dan 24.
9. Tahap 9 data ke 14 dan 26 bergabung menjadi kelompok 9, sehingga kelompok ini terdiri dari data ke 14, 28, 30, dan 26.
10. Tahap 10 data ke 6 dan 20 bergabung menjadi kelompok 10.
11. Tahap 11 data ke 14 dan 32 bergabung menjadi kelompok 11, sehingga kelompok ini terdiri dari data ke 14, 28, 30, 26, dan 32.
12. Tahap 12 data ke 3 dan 16 bergabung menjadi kelompok 12, sehingga data ini terdiri dari data ke 3, 16, dan 33.
13. Tahap 13 data ke 4 dan 17 bergabung menjadi kelompok 13.
14. Tahap 14 data ke 7 dan 21 bergabung menjadi kelompok 14, sehingga data ini terdiri dari data ke 7, 22, 21, dan 23.
15. Tahap 15 data ke 8 dan 14 bergabung menjadi kelompok 15, sehingga data ini terdiri dari data ke 8, 10, 24, 14, 28, 30, 26, dan 32.
16. Tahap 16 data ke 5 dan 18 bergabung menjadi kelompok 16.
17. Tahap 17 data ke 1 dan 11 bergabung menjadi kelompok 17.
18. Tahap 18 data ke 2 dan 19 bergabung menjadi kelompok 18.
19. Tahap 19 data ke 4 dan 12 bergabung menjadi kelompok 19, sehingga data ini terdiri dari data 4, 17, dan 12.

20. Tahap 20 data ke 3 dan 6 bergabung menjadi kelompok 20, sehingga data ini terdiri dari data ke 3, 16, 33, 6, dan 20.
21. Tahap 21 data ke 5 dan 9 bergabung menjadi kelompok 21, sehingga data ini terdiri dari data ke 5, 18, 9, dan 29.
22. Tahap 22 data ke 13 dan 31 bergabung menjadi kelompok 22.
23. Tahap 23 data ke 4 dan 7 bergabung menjadi kelompok 23, sehingga data ini terdiri dari data ke 4, 17, 12, 7, 22, 21, dan 23.
24. Tahap 24 data ke 1 dan 25 bergabung menjadi kelompok 24, sehingga data ini terdiri dari data ke 1, 11, dan 25.
25. Tahap 25 data ke 4 dan 8 bergabung menjadi kelompok 25.
26. Tahap 26 data ke 4 dan 5 bergabung menjadi kelompok 26.
27. Tahap 27 data ke 3 dan 15 bergabung menjadi kelompok 27.
28. Tahap 28 data ke 13 dan 34 bergabung menjadi kelompok 28, sehingga data ini terdiri dari data ke 13, 31, dan 34.
29. Tahap 29 data ke 1 dan 4 bergabung menjadi kelompok 29.
30. Tahap 30 data ke 2 dan 3 bergabung menjadi kelompok 30.
31. Tahap 31 data ke 2 dan 13 bergabung menjadi kelompok 31.
32. Tahap 32 data ke 1 dan 2 bergabung menjadi kelompok 32.
33. Tahap 33 data ke 1 dan 27 bergabung menjadi kelompok 33.

Tabel 4.8 Hasil Analisis *Cluster*

<i>Case</i>	<i>4 Clusters</i>	<i>3 Clusters</i>	<i>2 Clusters</i>
1	1	1	1
2	2	2	1
3	2	2	1
4	1	1	1
5	1	1	1
6	2	2	1
7	1	1	1
8	1	1	1
9	1	1	1
10	1	1	1
⋮	⋮	⋮	⋮
34	3	2	1

Adapun hasil *Cluster* tersebut sebagai berikut:

1. Jika 2 *Clusters*

*Cluster 1* : Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kepulauan Riau, Metro Jaya, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua.

*Cluster 2* : Sulawesi Selatan

2. Jika 3 *Clusters*

*Cluster 1* : Aceh, Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, Kepulauan Riau, Metro Jaya, Jawa Barat, Bali, NTB, Kalimantan Barat, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, dan Maluku Utara.

*Cluster 2* : Sumatera Utara, Sumatera Barat, Sumatera Selatan, Jawa Tengah, Jawa Timur, Banten, NTT, Kalimantan Tengah, Kalimantan Selatan, Maluku, Papua Barat dan Papua.

*Cluster 3* : Sulawesi Selatan

3. Jika 4 *Clusters*

*Cluster 1* : Aceh, Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, Kepulauan Riau, Metro Jaya, Jawa Barat, DI Yogyakarta, Bali, NTB, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, dan Maluku Utara.

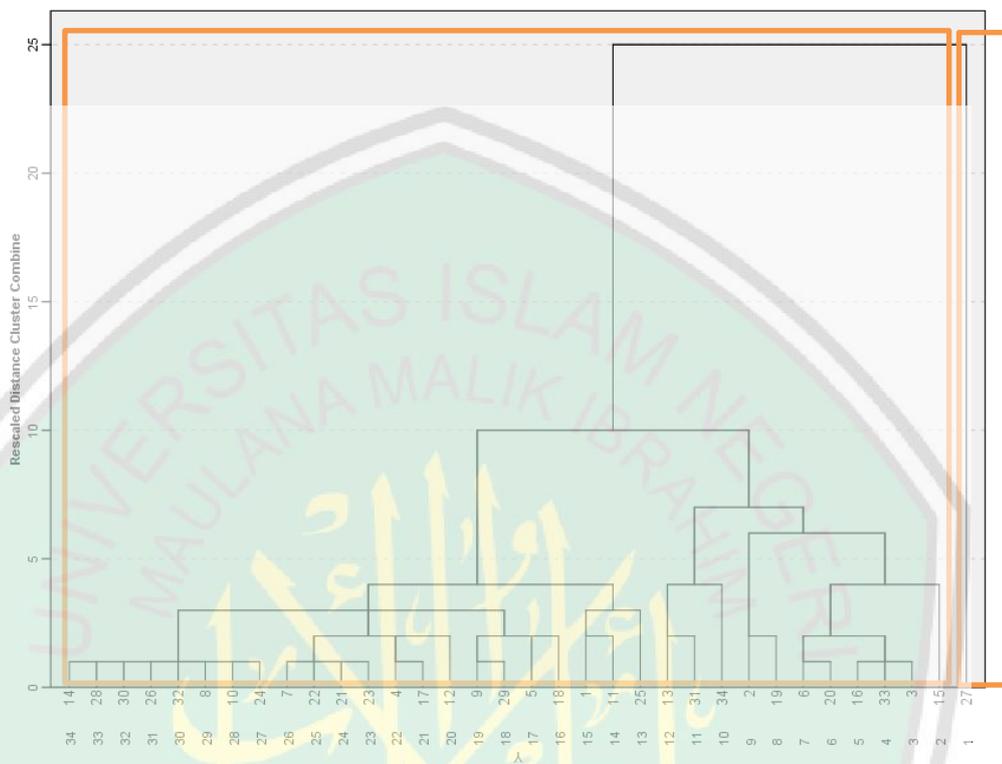
*Cluster 2* : Sumatera Utara, Sumatera Barat, Sumatera Selatan, Jawa Timur, Banten, NTT, Kalimantan Barat, dan Papua Barat.

*Cluster 3* : Jawa Tengah, Maluku dan Papua.

*Cluster 4* : Sulawesi Selatan



#### 4.2.2 Menentukan Jumlah Anggota *Cluster*



Gambar 4.2 Hasil Pemotongan Dendrogram

Proses pemotongan diatas berdasarkan pada jarak terjauh *cluster*. Sehingga dapat disimpulkan pada tabel 4.7 sebagai berikut.

Tabel 4.9 Anggota Masing-Masing *Cluster*

<i>Cluster</i>	Kepolisian Daerah
<i>Cluster 1</i>	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kepulauan Riau, Metro Jaya, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur,

	Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua
<i>Cluster 2</i>	Sulawesi Selatan

Pada tabel 4.9 dapat diketahui bahwa pengelompokan kepolisian daerah berdasarkan jumlah kriminalitas menjadi 2 *cluster* yang memiliki komposisi atau jumlah anggota yang berbeda diantara masing-masing cluster. Paling sedikit adalah *cluster 2* dengan hanya terdiri dari satu kepolisian daerah dan yang terbanyak adalah *cluster 1* yang terdiri dari 33 kepolisian daerah. Setelah menentukan jumlah dan anggota *cluster* maka langkah selanjutnya melakukan validasi dan profiling *cluster* dengan tujuan agar tiap-tiap *cluster* menghasilkan ciri khas masing-masing. Untuk tahap validasi dan profiling akan dijelaskan pada tabel 4.11 dan 4.12.

#### 4.3 Karakteristik Setiap Kelompok

Berdasarkan pengujian yang telah dilakukan menunjukkan bahwa terdapat perbedaan antara 2 kelompok yang terbentuk dengan metode *average linkage*. Setelah itu akan diketahui masing-masing kelompok berdasarkan variabel jumlah kriminalitas.

Tabel 4.10 Perbedaan Karakteristik Setiap Kelompok

Kelompok 1	Kelompok 2
- Mempekerjakan anak di bawah umur - penculikan	- Mempekerjakan anak dibawah umur - Korupsi

### 4.3.1 Uji Validitas dan Profiling *Cluster*

Tabel 4.11 *Cluster* 1

KELOMPOK 1			
Kepolisian Daerah	Mempekerjakan Anak dibawah Umur	Penculikan	Korupsi
A c e h	12	29	15
Sumatera Utara	73	15	29
Sumatera Barat	131	1	12
Riau	15	1	21
Jambi	49	4	19
Sumatera Selatan	129	3	22
⋮	⋮	⋮	⋮
Rata-rata	57.24242424	6.545454545	14.60606061
Max	234	51	38
Min	0	0	0

Tabel 4.12 *Cluster* 2

KELOMPOK 2			
Kepolisian Daerah	Mempekerjakan anak dibawah umur	Penculikan	Korupsi
Sulawesi Selatan	434	6	14
Rata-rata	434	6	14
Max	434	6	14
Min	434	6	14

Tujuan validasi *cluster* ini adalah untuk memastikan apakah kepolisian daerah tersebut masih menjadi sebuah kelompok ataukah terpisah jauh sehingga tidak dapat digolongkan sebagai sebuah kelompok lagi. Validasi ini dilakukan dengan membagi data menjadi dua kelompok besar dengan melakukan pemotongan *dendogram* menjadi dua kelas besar. Pembagian dua kelompok besar ini berdasarkan jarak terjauh yang bisa dilihat pada *dendogram*. Validitas *cluster* ini dilakukan tanpa mencantumkan iterasi jarak sebab fungsinya hanyalah untuk memastikan kekompakan pada *cluster* yang di bentuk pada langkah sebelumnya. Kedua kelompok besar tersebut akan dikelompokkan secara bergiliran. Dari masing-masing kelompok akan dibentuk *cluster* kembali dengan pemotongan *dendogram*.

*Cluster 1* terbentuk dari suatu gerombolan pada dendogram yang merupakan suatu hal yang wajar jika kepolisian daerah tersebut dapat dikatakan dalam satu kelompok. Jika dilihat dari interasinya kelompok ini di bentuk menurut tabel proses jarak. Dari tabel 4.7 dan 4.8 terlihat bahwa kota-kota tersebut diatas bergerombol menjadi satu. Proses pembentukan kelompok ini dimulai dengan pengelompokan antara Aceh dengan Sumatera Utara, Sumatera Barat dengan Riau, Jambi dengan Sumatera Selatan, Bengkulu dengan Lampung, Kepulauan Bangka Belitung dengan Kepulauan Riau, Metro Jaya dengan Jawa Barat, Jawa Tengah dengan DI Yogyakarta, Jawa Timur dengan Banten, Bali dengan Nusa Tenggara Barat, NTT dengan Kalimantan Barat, Kalimantan Tengah dengan Kalimantan Selatan, Kalimantan Timur dengan Kalimantan Utara, Sulawesi Utara dengan Sulawesi Tengah, Sulawesi Tenggara dengan Gorontalo, Sulawesi Barat dengan Maluku, Maluku Utara dengan Papua. Pada *cluster 2*, hanya terdiri dari Sulawesi Selatan. Rantai terakhir pada kelas ini adalah ketika dua kelompok tersebut membentuk rantai baru dengan kelompok lainnya.

Pada *cluster 1* memiliki nilai rata-rata kasus kriminalitas terbanyak yaitu mempekerjakan anak dibawah umur sebesar 57.24 kasus. Kasus penculikan sebesar 6.54 kasus dan kasus korupsi sebesar 14.61 kasus. Dengan kasus mempekerjakan anak dibawah umur yang paling dominan sebesar 234 kasus, penculikan 51 kasus dan korupsi 38 kasus dan paling sedikitnya tidak terdapat kasus. Pada *cluster 2* memiliki nilai rata-rata kasus kriminalitas terbanyak yaitu mempekerjakan anak dibawah umur sebesar 434 kasus. Kasus penculikan sebesar 6 kasus dan kasus korupsi sebesar 14 kasus. Dengan kasus mempekerjakan anak dibawah umur yang paling dominan sebesar 434 kasus, penculikan 6 kasus dan korupsi 14 kasus. Dari penjelasan tersebut dapat diketahui bahwa antara *cluster 1* dengan *cluster 2* memiliki heterogenitas dan homogenitas yang tinggi.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Dari penelitian yang telah dilakukan berdasarkan jumlah kriminalitas di Indonesia tahun 2019 dengan menggunakan analisis *cluster*. Dapat diambil kesimpulan sebagai berikut:

1. Jarak yang digunakan dalam analisis ini adalah:

$$d_{i,j} = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

Analisis *cluster* hirarki metode *average linkage* menggunakan persamaan:

$$d_{(IJ)K} = \frac{\sum a \sum b \sum ab}{N_{IJ}N_K}$$

2. Hasil pengelompokan analisis *cluster* hirarki metode *average linkage* sebanyak 2 *cluster*. Dengan terbaginya jenis kriminalitas menjadi sedikit *cluster*, hal ini menunjukkan bahwa jenis kriminalitas di Indonesia sangat homogen.
3. Variabel-variabel yang mencirikan setiap kelompok sebagai berikut:
  - a. Kelompok 1 dominan terhadap kasus mempekerjakan anak dibawah umur dan penculikan.
  - b. Kelompok 2 dominan terhadap kasus mempekerjakan anak dibawah umur dan korupsi.

#### 5.2 Saran

Saran untuk penelitian selanjutnya, penyang mencirikan eliti bisa mencoba melakukan analisis *cluster* menggunakan metode yang lain antara lain metode *ward* dan metode *centroid* yang penerapan metode tersebut dapat digunakan pada data indikator kesehatan, atau sebaran penyakit.

## DAFTAR RUJUKAN

- Anderberg MR. 1973. *Cluster Analysis for Application*. New York: Academic Press.
- Anggara, Mario dkk. 2016. *Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member di Alvaro Fitness*. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*. Vol.1, No.1.
- Bacher J, Wenzig K, VoglerM. 2004. *SPSS Two Step Cluster - A First Evaluation*. <http://www.statisticalinnovations.com/products/Two Step.pdf>.
- Baroroh, A. 2012. *Analisis Multivariat dan Time Series dengan SPSS 21*. Tangerang Selatan: PT. ELEK MEDIA KOMPUTINDO.
- Bluman, A. (2004). *Elementary Statistics: A Step By Step Approach, 5<sup>th</sup> Edition*. New York: Me Graw-Hill.
- BPS. 2019. *Statistik Kriminlitas 2019*. Jakarta: Badan Pusat Statistik.
- Gudono. 2011. *Analisis Data Multivariat*. Yogyakarta: BPFEE.
- Hair, J.F., Anderson R.E., Tatham, R.L., dan Black, W.C. 1995. *Multivariate Data Analysis with Readings*. New Jersey: Pearson New International Edition.
- Hocking, R. 1996. *Methods and Applications of Liniar Models: Regression and the Analysis Variance*. New York: John Wiley and Sons.
- Irianto, A. 2004. *Statistik Konsep Dasar & Apikasinya*. Jakarta: Prenada Media Group.
- Johnson, R.A & Wichem, DW. 1992. *Applied Multivariate Statistikal Analysis Third Edition*. New Jersey: Prentice Hall International.
- Kartono. 1999. *Patologi Sosial*. Jakarta: Raja grafindo Persada.

- Katsir, Ibnu Abu Al-Fida`. 1999. *Tafsir Al-Qur`an Al-A`zhim (Dar Thaybah)*.4/83.
- Machfudhoh, Siti dan Nuri Wahyuningsih. 2013. "Analisis Cluster Kabupaten/Kota Berdasarkan Pertumbuhan Ekonomi Jawa Timur". *Jurnal Sains Dan Seni Pomits*, Vol. 2, No.1
- Maulidina dan Setiawan, 2012. *Pengelompokan Kecamatan di Pulau Madura Berdasarkan sektor Pertanian sebelum dan setelah berdiri jembatan suramadu*. *Jurnal sains dan seni ITS*. 1(1) : 49-55
- Nugroho, S. 2008. *Statistik Multivariat Terapan*. Bengkulu : UNIB Press.
- Rachmatin, D. 2014. *Aplikasi MetodeMetode Agglomerative Dalam Analisis Klaster Pada Data Tingkat Polusi Udara*. *Jurnal Ilmiah Program Studi Matematika STKIP Siliwangi Bandung*, Vol 3, No. 2, September 2014.
- Santoso, S. 2002. *Buku Latihan SPSS Statistik Multivariat*. Jakarta : PT Elex Media Komputindo.
- Simamora B. 2005. *Analisis Multivariat Pemasaran*. PT. Gramedia Pustaka Utama: Jakarta.
- Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. Jakarta: Rineka Cipta.
- Supranto, J. 2010. *Analisis Multivariat Arti dan Interpretasi*. Cetakan Kedua. Rineka Cipta. Jakarta.

## LAMPIRAN-LAMPIRAN

Data

No	Kepolisihan Daerah	Mempekerjakan anak dibawah umur	Penculikan	Korupsi
1	A c e h	12	29	15
2	Sumatera Utara	73	15	29
3	Sumatera Barat	131	1	12
4	Riau	15	1	21
5	Jambi	49	4	19
6	Sumatera Selatan	129	3	22
7	Bengkulu	4	2	24
8	Lampung	2	5	4
9	Kep. Bangka Belitung	33	0	4
10	Kepulauan Riau	0	2	7
11	Metro Jaya	4	34	9
12	Jawa Barat	21	12	28
13	Jawa Tengah	192	7	29
14	DI Yogyakarta	13	1	3
15	Jawa Timur	149	12	35
16	Banten	136	0	5
17	Bali	26	3	21
18	Nusa Tenggara Barat	52	2	5
19	Nusa Tenggara Timur	72	6	18
20	Kalimantan Barat	128	1	29
21	Kalimantan Tengah	3	0	38
22	Kalimantan Selatan	2	0	23
23	Kalimantan Timur	0	0	35
24	Kalimantan Utara	1	0	1
25	Sulawesi Utara	0	51	5
26	Sulawesi Tengah	15	0	8
27	Sulawesi Selatan	434	6	14
28	Sulawesi Tenggara	11	1	3
29	Gorontalo	32	6	5
30	Sulawesi Barat	10	0	1
31	Maluku	193	12	4
32	Maluku Utara	18	2	0
33	Papua Barat	129	0	6
34	Papua	234	4	14

Matrik jarak *manhattan*

DATA	1	2	3	4	...	34
1	0.000	89.000	150.000	37.000	...	248.000
2	89.000	0.000	89.000	80.000	...	187.000
3	150.000	89.000	0.000	125.000	...	108.000
4	37.000	80.000	125.000	0.000	...	229.000
5	66.000	45.000	92.000	39.000	...	190.000
6	150.000	75.000	14.000	117.000	...	114.000
7	44.000	87.000	140.000	15.000	...	242.000
8	45.000	106.000	141.000	34.000	...	243.000
9	61.000	80.000	107.000	36.000	...	215.000
10	47.000	108.000	137.000	30.000	...	243.000
11	19.000	108.000	163.000	56.000	...	265.000
12	39.000	56.000	137.000	24.000	...	235.000
13	216.000	127.000	84.000	191.000	...	60.000
14	41.000	100.000	127.000	20.000	...	235.000
15	174.000	85.000	52.000	159.000	...	114.000
16	163.000	102.000	13.000	138.000	...	111.000
17	46.000	67.000	116.000	13.000	...	216.000
18	77.000	58.000	87.000	54.000	...	193.000
19	86.000	21.000	70.000	65.000	...	168.000
20	158.000	69.000	20.000	121.000	...	124.000
21	61.000	94.000	155.000	30.000	...	259.000
22	47.000	92.000	141.000	16.000	...	245.000
23	61.000	94.000	155.000	30.000	...	259.000
24	54.000	115.000	142.000	35.000	...	250.000
25	44.000	133.000	188.000	81.000	...	290.000
26	39.000	94.000	121.000	14.000	...	229.000
27	446.000	385.000	310.000	431.000	...	202.000
28	41.000	102.000	129.000	22.000	...	237.000
29	53.000	74.000	111.000	38.000	...	213.000
30	45.000	106.000	133.000	26.000	...	241.000
31	209.000	148.000	81.000	206.000	...	59.000
32	48.000	97.000	126.000	25.000	...	232.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
34	248.000	187.000	108.000	229.000	...	0.000

## Pengelompokan 1

DATA	1	2	3	4	...	34
1	0.000	89.000	150.000	37.000	...	248.000
2	89.000	0.000	89.000	80.000	...	187.000
3	150.000	89.000	0.000	125.000	...	108.000
4	37.000	80.000	125.000	0.000	...	229.000
5	66.000	45.000	92.000	39.000	...	190.000
6	150.000	75.000	14.000	117.000	...	114.000
7	44.000	87.000	140.000	15.000	...	242.000
8	45.000	106.000	141.000	34.000	...	243.000
9	61.000	80.000	107.000	36.000	...	215.000
10	47.000	108.000	137.000	30.000	...	243.000
11	19.000	108.000	163.000	56.000	...	265.000
12	39.000	56.000	137.000	24.000	...	235.000
13	216.000	127.000	84.000	191.000	...	60.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
34	248.000	187.000	108.000	229.000	...	0.000

## Pengelompokan 2

DATA	1	2	3	4	...	34
1	0.000	89.000	150.000	37.000	...	248.000
2	89.000	0.000	89.000	80.000	...	187.000
3	150.000	89.000	0.000	125.000	...	108.000
4	37.000	80.000	125.000	0.000	...	229.000
5	66.000	45.000	92.000	39.000	...	190.000
6	150.000	75.000	14.000	117.000	...	114.000
7	44.000	87.000	140.000	15.000	...	242.000
8	45.000	106.000	141.000	34.000	...	243.000
9	61.000	80.000	107.000	36.000	...	215.000
10	47.000	108.000	137.000	30.000	...	243.000
11	19.000	108.000	163.000	56.000	...	265.000
12	39.000	56.000	137.000	24.000	...	235.000
13	216.000	127.000	84.000	191.000	...	60.000
14,28	41.000	101.000	128.000	21.000	...	236.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
34	248.000	187.000	108.000	229.000	...	0.000

Pengelompokan 3

DATA	1	2	3	4	...	34
1	0.000	89.000	150.000	37.000	...	248.000
2	89.000	0.000	89.000	80.000	...	187.000
3	150.000	89.000	0.000	125.000	...	108.000
4	37.000	80.000	125.000	0.000	...	229.000
5	66.000	45.000	92.000	39.000	...	190.000
6	150.000	75.000	14.000	117.000	...	114.000
7	44.000	87.000	140.000	15.000	...	242.000
8	45.000	106.000	141.000	34.000	...	243.000
9	61.000	80.000	107.000	36.000	...	215.000
10	47.000	108.000	137.000	30.000	...	243.000
11	19.000	108.000	163.000	56.000	...	265.000
12	39.000	56.000	137.000	24.000	...	235.000
13	216.000	127.000	84.000	191.000	...	60.000
14,28,30	42.333	102.667	129.667	22.667	...	237.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
34	248.000	187.000	108.000	229.000	...	0.000

Pengelompokan 4

DATA	1	2	3	4	...	34
1	0.000	89.000	150.000	37.000	...	150.000
2	89.000	0.000	89.000	80.000	...	75.000
3	150.000	89.000	0.000	125.000	...	14.000
4	37.000	80.000	125.000	0.000	...	117.000
5	66.000	45.000	92.000	39.000	...	84.000
6	150.000	75.000	14.000	117.000	...	0.000
7,22	45.500	89.500	140.500	15.500	...	129.500
8	45.000	106.000	141.000	34.000	...	147.000
9	61.000	80.000	107.000	36.000	...	117.000
10	47.000	108.000	137.000	30.000	...	145.000
11	19.000	108.000	163.000	56.000	...	169.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	163.000	102.000	13.000	138.000	...	27.000

## Pengelompokan 5

DATA	1	2	3	4	...	34
∴	∴	∴	∴	∴	∴	∴
17	46.000	67.000	116.000	13.000	...	111.000
18	77.000	58.000	87.000	54.000	...	216.000
19	86.000	21.000	70.000	65.000	...	193.000
20	158.000	69.000	20.000	121.000	...	168.000
21,23	61.000	94.000	155.000	30.000	...	124.000
24	54.000	115.000	142.000	35.000	...	259.000
25	44.000	133.000	188.000	81.000	...	250.000
26	39.000	94.000	121.000	14.000	...	290.000
27	446.000	385.000	310.000	431.000	...	229.000
28	53.000	74.000	111.000	38.000	...	202.000
29	209.000	148.000	81.000	206.000	...	213.000
30	48.000	97.000	126.000	25.000	...	59.000
∴	∴	∴	∴	∴	∴	∴
34	111.000	216.000	193.000	168.000	...	0.000

## Pengelompokan 6

DATA	1	2	3	4	...	34
∴	∴	∴	∴	∴	∴	∴
10	47.000	108.000	137.000	30.000	...	243.000
11	19.000	108.000	163.000	56.000	...	265.000
12	39.000	56.000	137.000	24.000	...	235.000
13	216.000	127.000	84.000	191.000	...	60.000
14,28,30	42.333	102.667	129.667	22.667	...	237.667
15	174.000	85.000	52.000	159.000	...	114.000
16,33	159.000	98.000	11.000	134.000	...	114.000
17	46.000	67.000	116.000	13.000	...	216.000
18	77.000	58.000	87.000	54.000	...	193.000
19	86.000	21.000	70.000	65.000	...	168.000
20	158.000	69.000	20.000	121.000	...	124.000
21,23	61.000	94.000	155.000	30.000	...	259.000
24	54.000	115.000	142.000	35.000	...	250.000
∴	∴	∴	∴	∴	∴	∴
34	243.000	265.000	235.000	60.000	...	0.000

Pengelompokan 7

DATA	1	2	3	4	...	34
:	:	:	:	:	:	:
8	45.000	106.000	141.000	34.000	...	243.000
9,29	57.000	77.000	109.000	37.000	...	214.000
10	47.000	108.000	137.000	30.000	...	243.000
11	19.000	108.000	163.000	56.000	...	265.000
12	39.000	56.000	137.000	24.000	...	235.000
13	216.000	127.000	84.000	191.000	...	60.000
:	:	:	:	:	:	:
34	243.000	214.000	243.000	265.000	...	0.000

Pengelompokan ini dilanjutkan sampai pengelompokan ke 33

Pengelompokan 33

DATA	1,2,3,4,5,6,7,8,9,10, 11,1,2,13,14,15,16,17, 18,19,20,21,22,23,24,25,26, 28,29,30,31,32,33,34	27
1,2,3,4,5,6,7,8,9,10, 11,1,2,13,14,15,16,17, 18,19,20,21,22,23,24,25,26, 28,29,30,31,32,33,34	0.000	393.788
27	393.788	0.000

*Cluster 1*

KELOMPOK 1			
Kepolisian Daerah	Mempekerjakan Anak dibawah Umur	Penculikan	Korupsi
A c e h	12	29	15
Sumatera Utara	73	15	29
Sumatera Barat	131	1	12
Riau	15	1	21
Jambi	49	4	19
Sumatera Selatan	129	3	22
⋮	⋮	⋮	⋮
Rata-rata	57.24242424	6.545454545	14.60606061
Max	234	51	38
Min	0	0	0

*Cluster 2*

KELOMPOK 2			
Kepolisian Daerah	Mempekerjakan anak dibawah umur	Penculikan	Korupsi
Sulawesi Selatan	434	6	14
Rata-rata	434	6	14
Max	434	6	14
Min	434	6	14

## RIWAYAT HIDUP



Renta Windasari dilahirkan di Jombang pada tanggal 4 Desember 1997, anak tunggal dari pasangan Bapak Mochamad Rochim dan Ibu Jiati. Pendidikan dasar ditempuh di kampung halamannya di MI Al-Ilham Mejoyo Losari yang ditamatkan pada tahun 2010. Pada tahun yang sama dia melanjutkan pendidikan menengah pertama di SMP Negeri 1 Gudo Jombang. Pada tahun 2013 dia menamatkan pendidikannya, kemudian melanjutkan pendidikan menengah Kejurusan SMK Negeri Gudo Jombang dan menamatkan pendidikan tersebut pada tahun 2016. Pendidikan berikutnya dia tempuh di Universitas Islam Negeri Maulana Malik Ibrahim Malang melalui jalur SBMPTN Tulis dengan mengambil jurusan Matematika Fakultas Sains dan Teknologi.



**KEMENTERIAN AGAMA RI**  
**UNIVERSITAS ISLAM NEGERI**  
**MAULANA MALIK IBRAHIM MALANG**  
**FAKULTAS SAINS DAN TEKNOLOGI**  
**Jl. Gajayana No. 50 Dinoyo Malang Telp./Fax.(0341)558933**

**BUKTI KONSULTASI SKRIPSI**

Nama : Renta Windasari  
NIM : 16610058  
Fakultas/Jurusan : Sains dan Teknologi/Matematika  
Judul Skripsi : Analisis *Cluster* Hirarki Metode *Average Linkage*  
Berdasarkan Jumlah Kriminalitas di Indonesia Tahun 2019  
Pembimbing I : Ari kusumastuti, M.Si., M.Pd  
Pembimbing II : Muhammad Khudzaifah, M.Si

No	Tanggal	Hal	Tanda Tangan
1.	10 Februari 2020	Konsultasi Bab I & II	1.
2.	13 Februari 2020	Konsultasi Bab I & II Agama	2.
3.	18 Februari 2020	Revisi Bab I & II	3.
4.	25 Februari 2020	Konsultasi Bab III	4.
5.	13 Maret 2020	Revisi Bab I, Bab II & Bab III	5.
6.	31 Maret 2020	Konsultasi Bab IV	6.
7.	28 April 2020	Revisi Bab IV	7.
8.	30 April 2020	Konsultasi Seluruh Bab	8.
9.	29 Mei 2020	ACC Agama	9.
10.	2 Juni 2020	ACC Keseluruhan	10.

Malang, 8 Juni 2020  
Mengetahui,  
Ketua Jurusan Matematika

Dr. Usman Pagalay, M.Si  
NIP. 19650414 200312 1 001