

**KOMPARASI ANALISIS *K-MEDOIDS CLUSTERING*
DAN *HIERARCHICAL CLUSTERING***

(Studi Kasus : Data Kriminalitas di Indonesia Tahun 2020)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program
Studi Statistika



Disusun Oleh:

Shafa Bunga Faradilla

18611064

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA**

2022

**HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR**

Judul : Komparasi Analisis *K-Medoids Clustering* dan
Hierarchical Clustering (Studi Kasus : Data
Kriminalitas di Indonesia Tahun 2020)


Nama Mahasiswa : Shafa Bunga Faradilla


NIM : 18611064

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Mengetahui,
Ketua Prodi Statistika

Yogyakarta, 25 Maret 2022
Dosen Pembimbing Tugas Akhir


(Dr. Edy Widodo, M.Si.)


(Prof. Dr. Jaka Nugraha, S.Si., M.Si.)

HALAMAN PENGESAHAN
TUGAS AKHIR

**KOMPARASI ANALISIS *K-MEDOIDS CLUSTERING* DAN
*HIERARCHICAL CLUSTERING***

(Studi Kasus : Data Kriminalitas di Indonesia Tahun 2020)

Nama Mahasiswa : Shafa Bunga Faradilla

NIM : 18611064

**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL : 5 April 2022**

Nama Penguji

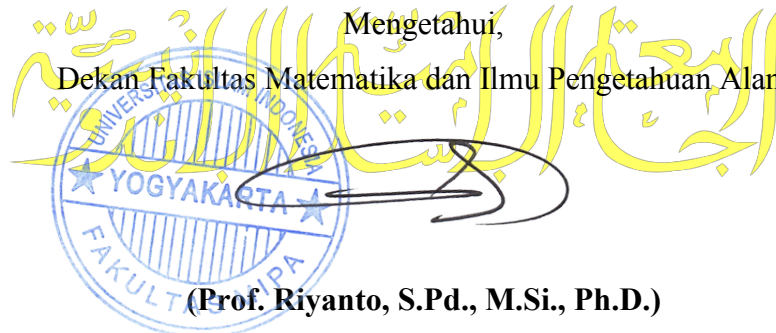
Tanda Tangan

1. Kariyam, S.Si., M.Si.
2. Dina Tri Utari, S.Si., M.Sc
3. Prof. Dr. Jaka Nugraha, S.Si.,M.Si.

.....
.....
.....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Prof. Riyanto, S.Pd., M.Si., Ph.D.)

KATA PENGANTAR



Assalamu'alaikum Wr.Wb

Alhamdulillah rabbil'alamin, puji syukur penulis ucapkan atas kehadiran Allah SWT yang telah melimpahkan rahmat serta karunianya, sehingga penulis dapat menyelesaikan Tugas Akhir ini. Shalawat serta salam sejahtera panjatkan kehadiran junjungan nabi besar Muhammad SAW yang mengantarkan manusia dari zama kegelapan menuju zaman yang terang benderang dan penuh dengan ilmu pengetahuan. Tugas Akhir ini disusun sebagai salah satu syarat untuk mendapatkan gelar Sarjana Statistika di Universitas Islam Indonesia. Tugas Akhir yang berjudul **“Komparasi Analisis *K-Medoids Clustering* dan *Hierarchical Clustering* (Studi Kasus: Data Tindakan Kriminalitas di Indonesia Tahun 2020)”**.

Selama proses penyusunan Tugas Akhir ini, tentunya tidak lepas dari bantuan dan dukungan dari beberapa pihak baik secara langsung maupun tidak langsung. Oleh karena itu pada kesempatan ini penulis bermaksud menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Yogyakarta.
2. Bapak Dr. Edy Widodo., M.Si., selaku Ketua Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
3. Bapak Prof. Dr. Jaka Nugraha, S.Si., M.Si, selaku Dosen Pembimbing Tugas Akhir yang telah memberikan bimbingannya selama penyusunan Laporan Tugas Akhir ini.
4. Seluruh Staff Pengajar Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia yang telah memberikan ilmu-ilmu yang bermanfaat bagi penulis.
5. Kedua orang tua serta seluruh keluarga besar yang selalu mencurahkan doa, dukungan, dan semangat untuk kelancaran segala urusan penulis.

6. Teman-teman satu bimbingan Tugas Akhir yang selalu kompak dan banyak memberikan bantuan dalam menyelesaikan Tugas Akhir ini.
7. Teman Statistika Angkatan 2018 yang telah menemani proses perjalanan penulis selama masa kuliah dalam mengembangkan diri dan memberikan bimbingan berharga bagi penulis.
8. Sahabat Cutak Grub, Legion, mbak fura, mbak dila, mbak ita, mbak safira yang selalu sabar mendengarkan keluh kesah selama menulis skripsi ini.
9. Kepada idol Korea yang selalu menemani disaat down dengan lagu-lagu yang bagus, Kim Namjoon, Kim Seokjin, Min Yoongi, Jung Hoseok, Park Jimin, Kim Taehyung, Jeon Jungkook, BTS.

Semoga Allah SWT senantiasa memberikan balasan setimpal kepada mereka semua. Penulis menyadari bahwa masih terdapat banyak kekurangan dalam penulisan Tugas Akhir ini dikarenakan keterbatasan pengetahuan dan kemampuan yang dimiliki oleh penulis. Oleh karena itu kritik dan saran yang bersifat membangun diharapkan demi kesempurnaan penulisan Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat bagi penulis dan khususnya pihak yang membutuhkan. Akhir kata, semoga Allah SWT selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua. Aamiin ya Rabbal'Alamin.

Wassalamualaikum Wr.Wb.

Yogyakarta, 25 Maret 2022



(Shafa Bunga Faradilla)

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR.....	ii
HALAMAN PENGESAHAN TUGAS AKHIR	iii
KATA PENGANTAR.....	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN	x
PERNYATAAN	xi
INTISARI	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA.....	6
2.1. Penelitian Terdahulu	7
BAB III LANDASAN TEORI.....	10
3.1. Kriminalitas.....	10
3.2. Asumsi Analisis <i>Cluster</i>	11
3.2.1 Sampel Representatif.....	11
3.2.2 Multikolinieritas	12
3.3. <i>Principal Component Analysis</i> (PCA).....	12
3.4. Penentuan Jumlah <i>Cluster</i> Optimum	14
3.5. Analisis <i>Cluster</i>	14
3.6. <i>Non Hierarchical Methods</i> (Metode <i>Non Hierarchical</i>).....	15
3.6.1 <i>Clustering K-Medoids</i>	15
3.6.2 <i>Multikolinieritas</i>	12
3.7. <i>Hierarchical Methods</i> (Metode <i>Hierarchical</i>).....	16
BAB IV METODOLOGI PENELITIAN	18
4.1. Populasi Penelitian	18
4.2. Metode Pengambilan Data	18
4.3. Variabel penelitian	18
4.4. Metode Analisis Data	19
BAB V HASIL DAN PEMBAHASAN.....	20
5.1. Analisis Deskriptif	20
5.2. Uji Asumsi <i>Cluster</i>	26
5.2.1 Uji <i>Kaiser-Meyer Olkin</i>	26
5.2.2 Uji Multikolinieritas	26
5.3. <i>Principal Component Analysis</i> (PCA)	27
5.4. Penentuan Jumlah <i>Cluster</i> Optimum	29
5.5. <i>K-Medoids Clustering</i>	29
5.6. <i>Clustering Hierarchical</i>	31
5.7. Pemilihan Metode Terbaik.....	33

5.8. Interpretasi Metode Terbaik.....	34
BAB VI PENUTUP	36
6.1. Kesimpulan	36
6.2. Saran.....	36
DAFTAR PUSTAKA.....	38
LAMPIRAN	41

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Sebelumnya	7
Tabel 4.1 Variabel Penelitian	18
Tabel 5.1 Koefisien Korelasi	26
Tabel 5.2 Koefisien Korelasi	26
Tabel 5.3 Tabel <i>Eigenvalues</i>	27
Tabel 5.4 Hasil PCA	27
Tabel 5.5 Penentuan Jumlah <i>Cluster</i> Optimum	29
Tabel 5.6 Hasil <i>K-Medoids Clustering</i>	30
Tabel 5.7 <i>Agglomerative Coefficient</i>	31
Tabel 5.8 Hasil Pengelompokkan <i>Ward</i>	32
Tabel 5.9 Hasil Perbandingan Metode Terbaik	34
Tabel 5.10 Profilisasi <i>Cluster</i> Terbaik	34

DAFTAR GAMBAR

Gambar 4.1 <i>Flowchart</i>	19
Gambar 5.1 Kejahatan Terhadap Nyawa.....	20
Gambar 5.2 Kejahatan Terhadap Fisik.....	21
Gambar 5.3 Kejahatan Terhadap Kesusilaan	21
Gambar 5.4 Kejahatan Terhadap Kemerdekaan Orang.....	22
Gambar 5.5 Kejahatan Terhadap Penggunaan Kekerasan	23
Gambar 5.6 Kejahatan Terhadap Tanpa Penggunaan Kekerasan	23
Gambar 5.7 Kejahatan Terkait Narkotika	24
Gambar 5.8 Kejahatan Terhadap Penipuan, Penggelapan, Korupsi.....	25
Gambar 5.9 Grafik <i>K-Medoids Clustering</i>	30
Gambar 5.10 Pemetaan Kelompok Provinsi	31
Gambar 5.11 Dendrogram <i>Ward</i>	32
Gambar 5.12 Pemetaan Pengelompokkan Provinsi	33

DAFTAR LAMPIRAN

Lampiran 1	41
Lampiran 2	44

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 25 Maret 2022



(Shafa Bunga Faradilla)

INTISARI

KOMPARASI ANALISIS *K-MEDOIDS CLUSTERING* DAN *HIERARCHICAL CLUSTERING*

(Studi Kasus : Data Kriminalitas di Indonesia Tahun 2020)

Shafa Bunga Faradilla

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Kriminalitas adalah suatu perbuatan atau tindakan kejahatan yang dapat menimbulkan masalah-masalah dan mengakibatkan dampak negatif bagi kehidupan masyarakat. Di Indonesia kasus kriminalitas mengalami kenaikan di tiap tahunnya. Dengan banyaknya kasus kriminalitas di Indonesia, peneliti tertarik untuk membuat proses peng*clusteran* terhadap kasus kriminalitas di Indonesia tahun 2020 yang bertujuan agar masyarakat waspada terhadap kejahatan dan pemerintah lebih memperhatikan dan memperketat sistem keamanan untuk daerah tersebut. Analisis *clustering* merupakan analisis yang bertujuan untuk penempatan sekumpulan obyek dalam dua atau lebih kelompok berdasarkan kesamaan objek karakteristiknya. Hasil perbandingan menggunakan validasi internal (indeks *connectivity*, *dunn*, *silhouette*) didapatkan hasil terbaik yaitu algoritma *Hierarchical Clustering* metode *Ward* dengan diperoleh 2 *cluster* yaitu *cluster* 1 terdiri dari 29 provinsi dan *cluster* 2 terdiri dari 5 provinsi.

Kata Kunci : Kriminalitas, *K-Medoids Clustering*, *Hierarchical Clustering*.

ABSTRACT

COMPARISON ANALYSIS OF K-MEDOIDS *CLUSTERING* AND HIERARCHICAL *CLUSTERING*

(Case Study : Crime Data in Indonesia in 2020)

Shafa Bunga Faradilla

Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Islam Indonesia

Crime is an act or crime that can cause problems and have a negative impact on people's lives. In Indonesia, crime cases have increased every year. With so many criminal cases in Indonesia, researchers are interested in making a clustering process for criminal cases in Indonesia in 2020 which aims to make the public aware of crime and the government to pay more attention to and tighten the security system for the area. Clustering analysis is an analysis that aims to place a set of objects in two or more groups based on the similarity of their characteristic objects. The results of the comparison using internal validation (connectivity index, dunn, silhouette) obtained the best results, namely the Hierarchical Clustering algorithm Ward method with 2 clusters obtained, namely cluster 1 consisting of 29 provinces and cluster 2 consisting of 5 provinces.

Keywords: *Crime, K-Medoids Clustering, Hierarchical Clustering*

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Salah satu faktor pemicu terjadinya tindakan kriminalitas yaitu masalah perekonomian serta kesadaran masyarakat untuk melaporkan tindakan kriminalitas. Banyaknya jumlah penduduk menimbulkan berbagai masalah sosial di masyarakat seperti kemiskinan, pengangguran, dan lainnya. Adanya masalah sosial tersebut mendorong beberapa orang melakukan tindakan kriminalitas.

Kriminalitas adalah suatu perbuatan atau tindakan kejahatan yang dapat menimbulkan masalah-masalah dan mengakibatkan dampak negatif bagi kehidupan masyarakat. Seringkali, tindakan ini akan merugikan banyak pihak dan pelaku tindakannya disebut sebagai seorang kriminal. Kriminalitas sering terjadi di negara lain, termasuk Indonesia. Jenis-jenis kriminalitas yang sering terjadi di Indonesia adalah kejahatan terhadap nyawa (pembunuhan), kejahatan terhadap fisik (penganiyaan), kejahatan terhadap kesusilaan (pemeriksaan), pencurian, narkoba, penipuan bahkan korupsi (Supratman, 2020). Berdasarkan hasil survey tingkat kriminalitas, Indonesia menduduki peringkat keempat di ASEAN pada tahun 2020 (Asia, 2020).

Rasa aman merupakan salah satu hak asasi yang harus diperoleh setiap orang seperti yang disebutkan pada pasal 28G ayat 1 dalam UUD Republik Indonesia 1945 yang berbunyi “Setiap orang berhak atas perlindungan diri pribadi, keluarga, kehormatan, martabat, dan harta benda yang di bawah kekuasaannya, serta berhak atas rasa aman dan perlindungan dari ancaman ketakutan untuk berbuat atau tidak berbuat sesuatu yang merupakan hak asasi”.

Jumlah kejadian kejahatan atau tindakan kejahatan pada periode 2018-2020 di Indonesia mengalami cenderung fluktuatif. Menurut Badan Pusat Statistik, jumlah kejadian kejahatan di tahun 2018 sebanyak 294.281 kejadian, lalu terjadi penurunan pada tahun 2019 sebanyak 269.324 kejadian, dan menurun pada tahun 2020 sebanyak 247.218 kejadian. Dari jumlah kejahatan menurut Polda/Provinsi pada tahun 2020, Polda Sumatera Utara merupakan jumlah kejadian tertinggi sebanyak 32.990 kejadian, diikuti dengan Polda Metro Jaya (DKI Jakarta) sebanyak

26.585 kejadian dan Polda Jawa Timur sebanyak 17.642 kejadian. Sementara itu, tiga wilayah dengan jumlah kejadian paling sedikit dengan jumlah kejadian kejahatan berturut-turut yaitu Polda Sulawesi Barat, Kalimantan Utara dan Maluku Utara dengan jumlah kejadian sebanyak 1.704, 1.015, dan 850 kejadian.

Indikator angka yang dapat menunjukkan tingkat kejahatan suatu wilayah dalam waktu tertentu disebut dengan *crime rate*. Semakin tinggi *crime rate* maka tingkat kerawanan akan kejahatan suatu daerah semakin tinggi juga, dan sebaliknya. Selama periode 2018-2020, tingkat resiko terkena kejahatan mengalami penurunan dan interval waktu terjadi kejadian tindak kejahatan yang semakin menurun. Indikator kriminalitas lainnya yaitu *crime clock* yang menunjukkan selang waktu terjadinya tindakan kejahatan. *Crime clock* pada periode 2018-2020 menunjukkan pola perkembangan yang serupa. Pada tahun 2018 selang waktu yang terjadi yaitu 1 menit 47 detik dan mengalami kenaikan pada tahun 2019 menunjukkan selang waktu sebesar 1 menit 57 detik. Kemudian, interval waktunya semakin panjang pada tahun 2020 yaitu 2 menit 7 detik. Interval waktu yang semakin panjang menunjukkan intensitas kejadian tindak kejahatan yang semakin menurun dan sebaliknya.

Dari banyaknya jumlah masing-masing kejahatan, terdapat beberapa contoh kasus kriminalitas di Indonesia, yaitu dilansir dari Kompas.com kasus narkoba di tahun 2020 dilaporkan oleh Polres Jakarta Barat penangkapan artis Lucinta Luna pada tanggal 11 Februari 2020, dari penangkapan tersebut polisi menyita obat penenang berjenis tramadol dan riklona yang dikategorikan dalam kategori psikotropika (Arby, 2020). Dari kasus ini maka menambah jumlah kejahatan terkait narkoba yang ada di Jakarta. Selain itu juga, terdapat kasus kejahatan lainnya yaitu kejahatan pembunuhan di Sumatera Utara berdasarkan berita inews.id polda Sumatera Utara mengungkapkan kasus pembunuhan Hakim Pengadilan Negeri Medan (Jamaludin) pada awal Januari 2020 (Rasyid, Purba, & A, 2020).

Dengan banyaknya tindakan kejahatan atau kriminalitas maka perlu adanya pengclusteran daerah-daerah rawan kriminalitas menurut Polda/Provinsi di Indonesia, agar masyarakat waspada terhadap kejahatan dan pemerintah lebih memperhatikan dan memperketat sistem keamanan untuk daerah tersebut. Oleh karena itu, maka dibuatlah pengolahan data yang dapat menyelesaikan

permasalahan tersebut yaitu pengelompokan daerah rawan tindak kejahatan atau kriminalitas menurut Polda/Provinsi di Indonesia menggunakan analisis *cluster*. Analisis *cluster* mempunyai tujuan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengklasifikasi objek sehingga objek-objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama. (Ediyanto, Muhlahsal, & Neva, 2013).

Diantara banyaknya *cluster*, metode pengclusteran dalam analisis *cluster* ada 2, yaitu metode *hierarchical* dan metode *non hierarchical*. Analisis *cluster* dengan metode *hierarchical* adalah analisis yang pengclusteran datanya dilakukan dengan cara mengukur jarak kedekatan pada setiap objek yang kemudian membentuk sebuah dendrogram. Terdapat beberapa metode dalam hierarki diantaranya yaitu, metode *single linkage*, *complete linkage*, *average linkage*, dan *ward*. Berbeda dengan metode *hierarchical*, metode *non hierarchical* dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan. Setelah jumlah *cluster* diketahui baru proses *cluster* dilakukan. Dalam metode *non hierarchical* terdapat 2 cara analisis yaitu *K-means Clustering* dan *K-Medoids Clustering*. Dalam penelitian ini menggunakan *K-Medoids Clustering* dikenal sebagai *Partitioning Around Medoids (PAM)* adalah varian dari metode *k-means*. *K-Medoids* hadir untuk mengatasi kelemahan *k-means clustering* yang sensitive terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data.

Dengan banyaknya kasus kriminalitas di Indonesia, maka perlu dilakukan pengelompokan daerah rawan kekerasan. Maka dari itu peneliti melakukan penelitian tentang *Clustering* yang mana metode yang digunakan adalah *K-Medoids Clustering* dan *Hierarchical Clustering* di Indonesia dengan data jumlah kriminalitas berdasarkan bentuknya tahun 2020 untuk menentukan menentukan *clustering* terbaik. Analisis *clustering* bertujuan untuk mencari pola data yang memiliki kemiripan sehingga terjadi beberapa kemungkinan dalam mengelompokkan data-data yang mirip tersebut.

1.2. Rumusan Masalah

Berdasarkan pemaparan pada latar belakang di atas, maka rumusan masalah yang akan dibahas pada penelitian ini adalah:

1. Bagaimana analisis deskriptif dari jumlah jumlah tindakan kriminalitas di Indonesia tahun 2020?
2. Bagaimana hasil *clustering* dari algoritma *K-medoids* pada jumlah tindakan kriminalitas di Indonesia tahun 2020?
3. Bagaimana hasil *clustering* dari algoritma *Hieararchical Clustering* pada jumlah jumlah tindakan kriminalitas di Indonesia tahun 2020?
4. Bagaimana hasil perbandingan *clustering* terbaik antara algoritma *K-medoids* dan *Hierarchical Clustering* dengan validasi *cluster* pada jumlah tindakan kriminalitas di Indonesia tahun 2020?

1.3. Batasan Masalah

Batasan masalah yang digunakan untuk memberi batasan beberapa masalah yang akan diangkat dan tidak menyimpang dari masalah penelitian. Dalam penelitian ini yang menjadi batasan masalah sebagai berikut:

1. Penelitian hanya berfokus pada jumlah kasus jumlah tindakan kriminalitas di Indonesia tahun 2020.
2. Data yang digunakan adalah data sekunder yang diperoleh dari Badan Pusat Statistik.
3. Metode analisis yang digunakan adalah analisis deskriptif, *K-Medoids Clustering* dan *Hieararchical Clustering*.
4. Pengolahan data menggunakan bantuan *software R Studio, Ms.Excel*.

1.4. Tujuan Penelitian

Adapun tujuan penelitian berdasarkan rumusan masalah di atas sebagai berikut:

1. Mengetahui analisis deskriptif dari jumlah tindakan kriminalitas di Indonesia tahun 2020.
2. Mengetahui hasil *clustering* dari algoritma *K-medoids* pada jumlah tindakan kriminalitas di Indonesia tahun 2020.
3. Mengetahui hasil *clustering* dari algoritma *Hieararchical Clustering* pada jumlah tindakan kriminalitas di Indonesia tahun 2020.
4. Mengetahui hasil perbandingan *clustering* terbaik antara algoritma *K-medoids* dan *Hierarchical Clustering* dengan validasi *cluster* pada jumlah tindakan kriminalitas di Indonesia tahun 2020.

1.5. Manfaat Penelitian

Adapun manfaat penelitian yang diperoleh dari penelitian ini sebagai berikut:

1. Diharapkan pihak kepolisian khususnya Kepolisian Republik Indonesia dapat mengklasifikasikan provinsi yang rawan kriminalitas dan memungkinkan untuk pihak Kepolisian dalam hal itu Polda, Polres, atau Polsek diprovinsi tersebut untuk lebih meningkatkan keamanan, pencegahan dan juga pemberantasan terhadap kriminalitas tersebut agar masyarakat lebih waspada lagi terhadap potensi kriminalitas yang dapat terjadi diwilayah masing- masing.
2. Bagi peneliti guna menjadi tingkat kewaspadaan terhadap diri sendiri.

BAB II

TINJAUAN PUSTAKA

Dalam penelitian diperlukannya dukungan hasil-hasil penelitian sebelumnya guna menjadi tolak ukur dalam meneliti. Penelitian yang dilakukan oleh M. Fariz Fadillah Mardianto,dkk pada tahun 2015 dengan studi kasus “Pengelompokkan Kelompok Optimal Kabupaten dan Kota Rawan Kriminalitas di Jawa Timur dengan Metode Analisis Kluster Terbaik”. Variabel penelitian yang digunakan terdapat 11 variabel yaitu yang terdiri dari jenis-jenis bentuk tindakan kriminalitas dan menggunakan metode *non-hierarchical* yaitu dengan *k-means* dan hierarki dengan menggunakan *single linkage*, *complete linkage*, *average linkage* dan metode *ward*. Hasil penelitian diperoleh pengelompokkan dengan hasil optimal yaitu terdapat 5 kelompok berdasarkan metode *ward* (Mardianto & dkk, 2015).

Dari penelitian yang dilakukan oleh Simatupang dan Wijayanto (2019) melakukan pengelompokkan tentang apa dan obejknya dengan hasil. Penelitian ini menggunakan variabel jenis-jenis kejahatan dengan menggunakan metode analisis *cluster K-Means dan Fuzzy C-Means*. Penentuan jumlah *cluster* terbaik dengan metode Elbow dapat menghasilkan jumlah *cluster* sebanyak tiga. Hasil penentuan jumlah *cluster* terbaik dengan metode *Elbow* akan dijadikan default untuk proses karakteristik berdasarkan studi kasus yang dilakukan. Dengan melihat *dunn index*, maka metode *clustering* yang terbaik adalah *fuzzy c-means*. Namun, apabila dilihat dari *connectivity* dan *silhouette*, metode *clustering* yang terbaik adalah *k-means*.

Penelitian yang dilakukan oleh Milla Alifatun Nahdliyah,dkk pada tahun 2019 dengan studi kasus “Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018”. Penelitian ini menggunakan metode “Metode *K-Medoids Clustering* dengan Validasi *Silhouette Index* dan *C-Index*”. Hasil penelitian ini adalah Pengclusteran menggunakan metode *k-medoids* dengan jarak *euclidean* dan *manhattan* untuk $k = 3, 4, \text{ dan } 5$ diperoleh *cluster* yang optimal pada $k = 4$ dengan jarak *euclidean* dimana nilai $SI = 0,3862593$ dan $CI = 0,043893$. Berdasarkan hasil pengclusteran pada metode ini didapatkan bahwa jarak pengukuran yang digunakan akan berpengaruh terhadap hasil pengclusteran (Nahdliyah & dkk, 2019).

Penelitian yang dilakukan oleh Anggi Dwi Lestari, dkk pada tahun 2019 dengan studi kasus “Pengelompokan Provinsi di Indonesia berdasarkan kriminalitas menggunakan metode *Ward* dan *K-medoids*”. Variabel yang digunakan adalah kejahatan terhadap nyawa, kejahatan fisik/badan, kejahatan terhadap keasusilaan, kejahatan terhadap kemerdekaan orang, kejahatan terhadap hak milik/barang dengan penggunaan kekerasan, kejahatan terhadap hak milik/barang tanpa penggunaan kekerasan, kejahatan terkait narkoba, dan kejahatan terkait penipuan, penggelapan dan korupsi dari 34 Provinsi di Indonesia. Hasil penelitian yang diperoleh yaitu terdapat 4 *cluster* dan metode *ward* merupakan metode terbaik dibandingkan *k-medoids* karena dilihat dari nilai rasio terkecil yang didapat bahwa nilai rasio yaitu 0.276368 (Lestari & dkk).

Penelitian yang digunakan oleh Hotma Dame Tampubolon, dkk pada tahun 2021 dengan studi kasus “Penerapan Algoritma *K-Means* dan *K-Medoids Clustering* untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi”. Data yang digunakan adalah tindakan kriminalitas tahun 2017-2019 diantaranya yaitu tindak kejahatan terhadap nyawa, kejahatan terhadap fisik, kejahatan terhadap kesusilaan, kejahatan terhadap hak milik/barang, narkoba, serta penipuan dan korupsi. Hasil penelitian yang diperoleh yaitu pada algoritma *k-means* terdapat 2 *cluster* yaitu dikategorika pada *cluster* 1 merupakan tindak kriminalitas tinggi, dan *cluster* 2 merupakan tindak kriminalitas rendah (Tampubolon & dkk, 2021).

2.1. Penelitian Terdahulu

Tabel 2.1 Tabel Penelitian Sebelumnya

Peneliti (Tahun)	Judul	Metode	Studi Kasus	Hasil Penelitian
M. Fariz Fadillah Mardianto, dkk (2015)	Pengelompokan Kelompok Optimal Kabupaten dan Kota Rawan Kriminalitas di Jawa Timur dengan Metode Analisis Kluster Terbaik	Metode <i>Ward</i>	Data kota rawan kriminalitas di Jawa Timur	Hasil penelitian diperoleh pengelompokan dengan hasil optimal yaitu terdapat 5 kelompok berdasarkan metode <i>ward</i>
Simatupang dan wijayanto (2019)	Analisis <i>Cluster</i> berdasarkan tindakan kriminalitas di	Metode Elbow	Data kriminalitas di Indonesia tahun 2019	Penentuan jumlah <i>cluster</i> terbaik dengan metode Elbow dapat menghasilkan jumlah

	Indonesia tahun 2019			<i>cluster</i> sebanyak tiga. Hasil penentuan jumlah <i>cluster</i> terbaik dengan metode <i>Elbow</i> akan dijadikan default untuk proses karakteristik berdasarkan studi kasus yang dilakukan. Dengan melihat <i>dunn index</i> , maka metode <i>clustering</i> yang terbaik adalah <i>fuzzy c-means</i> . Namun, apabila dilihat dari <i>connectivity</i> dan <i>silhouette</i> , metode <i>clustering</i> yang terbaik adalah k-means
Milla Alifatun Nahdliyah, dkk (2019)	Metode <i>K-Medoids Clustering</i> dengan Validasi <i>Silhouette Index</i> dan <i>C-Index</i>	Metode <i>K-Medoids Clustering</i> dengan Validasi <i>Silhouette Index</i> dan <i>C-Index</i>	Jumlah kriminalitas Kabupaten/kota di Jawa Tengah Tahun 2018	Pengclusteran menggunakan metode <i>k-medoids</i> dengan jarak <i>euclidean</i> dan <i>manhattan</i> untuk $k = 3, 4,$ dan 5 diperoleh <i>cluster</i> yang optimal pada $k = 4$ dengan jarak <i>euclidean</i> dimana nilai $SI = 0,3862593$ dan $CI = 0,043893$. Berdasarkan hasil pengclusteran pada metode ini didapatkan bahwa jarak pengukuran yang digunakan akan berpengaruh terhadap hasil pengclusteran
Anggi Dwi Lestari, dkk (2019)	Pengelompokan Provinsi di Indonesia berdasarkan kriminalitas menggunakan metode <i>Ward</i> dan <i>K-medoids</i>	Metode <i>Ward</i> dan <i>K-Medoids</i>	Data kriminalitas di Indonesia tahun 2018	Diperoleh yaitu terdapat 4 <i>cluster</i> dan metode <i>ward</i> merupakan metode terbaik dibandingkan <i>k-medoids</i> karena dilihat dari nilai rasio terkecil yang didapat bahwa nilai rasio yaitu 0.276368
Hotma Dame Tampubolon, dkk (2021)	Penerapan Algoritma <i>K-Means</i> dan <i>K-Medoids Clustering</i> untuk Mengelompokkan Tindak	Metode <i>K-Means</i> dan <i>K-Medoids Clustering</i>	Data kriminalitas menurut Provinsi tahun 2017-2019	Hasil penelitian yang diperoleh yaitu pada algoritma k-means terdapat 2 <i>cluster</i> yaitu dikategorika pada <i>cluster</i> 1 merupakan tindak kriminalitas tinggi, dan

	Kriminalitas Berdasarkan Provinsi			<i>cluster</i> 2 merupakan tindak kriminalitas rendah
--	-----------------------------------	--	--	---

Penelitian-penelitian sebagaimana terlihat pada tabel 2.1 digunakan sebagai acuan dalam penelitian ini dengan Sebagian besar penelitian memiliki tema “Analisis *Clustering* Kriminalitas Indonesia”. Terdapat perbedaan dari penelitian sebelumnya yaitu penelitian ini mengkombinasikan antara metode *non hierarchical* dan *hierarchical*, variabel yang digunakan dengan penelitian sebelumnya berbeda, dan periode yang digunakan dengan penelitian sebelumnya berbeda.

BAB III

LANDASAN TEORI

3.1. Kriminalitas

Kriminalitas adalah suatu perbuatan yang menyimpang, merugikan orang lain dan melanggar hukum, agama serta norma-norma sosial (Dewi, Windarto, Damanik, & H, 2019). Kriminalitas merupakan salah satu permasalahan yang sering terjadi di lingkungan masyarakat yang perlu diperhatikan karena merugikan berbagai kepentingan dan menimbulkan dampak negatif kepada masyarakat seperti kecemasan, rasa tidak aman, kepanikan, dan juga ketakutan. Kriminalitas dibagi menjadi beberapa jenis, sebagai berikut:

1. Kejahatan Terhadap Nyawa

Menurut Badan Pusat Statistik (BPS) kejahatan terhadap nyawa merupakan salah satu kejahatan yang paling tinggi hierarkinya dalam klasifikasi kejahatan internasional, selain itu dari sisi hukuman juga yang paling berat hukumannya dalam KUHP Indonesia. Kejahatan terhadap nyawa meliputi tindakan pembunuhan.

2. Kejahatan Terhadap Fisik

Menurut Badan Pusat Statistik (BPS) kejahatan fisik termasuk didalamnya adalah penganiyaan ringan, penganiyaan berat, dan kekerasan dalam rumah tangga. Kejahatan fisik adalah perbuatan yang mengakibatkan rasa sakit, jatuh sakit atau luka berat dalam kaitannya dengan akibat kejahatan, yaitu mengakibatkan penyakit (luka/cacat/sakit) atau halangan untuk menjalankan pekerjaan sehari-hari (luka berat) maupun tidak. Pemukulan, penamparan, pengeroyokan, termasuk kategori tindakan kejahatan terhadap fisik.

3. Kejahatan Terhadap Kesusialaan

Menurut Badan Pusat Statistik (BPS) klasifikasi kejahatan terhadap kesusialaan adalah jenis kejahatan pencabulan dan perkosaan. Pelecehan seksual adalah setiap perbuatan merendahkan, menghina, menyerang, dan/atau perbuatan lainnya terdapat tubuh, Hasrat seksual seseorang, dan/ atau fungsi reproduksi secara paksa.

4. Kejahatan Terhadap Kemerdekaan Orang

Kejahatan terhadap orang meliputi jenis kejahatan penculikan dan mempekerjakan anak di bawah umur. Kejahatan terhadap orang adalah kejahatan yang berkenaan dengan hak asasi manusia atau hak seseorang untuk bebas menggerakkan badan memenuhi kepentingan dalam masyarakat.

5. Kejahatan Terhadap Hak/Milik dengan Penggunaan Kekerasan

Jenis kejahatan ini termasuk dalam pencurian dengan kekerasan yang menggunakan senjata api dan pencurian dengan kekerasan menggunakan senjata tajam.

6. Kejahatan Terhadap Hak/Milik dengan Tanpa Penggunaan Kekerasan

Klasifikasi kejahatan ini termasuk dalam pencurian dengan pemberatan, pencurian kendaraan bermotor, pengrusakan/penghancuran barang, pembakaran dengan sengaja, dan penadahan.

7. Kejahatan Terkait Narkotika

Klasifikasi kejahatan ini adalah yang menggunakan narkotika dan psikotropika. Kejahatan narkotika merupakan suatu kejadian transasional yang saat ini menjadi suatu kejahatan lintas negara yang terjadi di Indonesia.

8. Kejahatan Terkait Penipuan, Penggelapan, dan Korupsi

Kejahatan ini memiliki tindak pidana penggelapan diatur dalam pasal 372 KUHP yang berbunyi “Barang siapa dengan sengaja menguasai secara melawan hukum sesuatu benda yang seharusnya atau sebagian merupakan kepunyaan orang lain yang berada padanya bukan karena kejahatan, karena bersalah melakukan penggelapan, dipidana dengan pidana penjara selama-lamanya 4 tahun”.

3.2. Asumsi Analisis *Cluster*

Terdapat dua asumsi yang harus dipenuhi dalam analisis *cluster* yaitu sebagai berikut.

3.2.1 Sampel Representatif

Sampel representatif merupakan sampel yang karakteristiknya sama dengan populasi. Penggunaan sampel yang representatif akan memberikan hasil maksimal dan sesuai dengan kondisi populasi yang ada. Jika penelitian menggunakan data populasi, maka dapat disimpulkan bahwa asumsi representatif terpenuhi (Hair, 1998).

Terdapat cara untuk melihat apakah sampel representatif atau tidak yaitu dengan menggunakan uji Kaiser-Meyer Olkin (KMO). Uji KMO digunakan untuk mengukur kecukupan pengambilan sampel secara keseluruhan dan mengukur kecukupan pengambilan sampel untuk setiap indikator. Apabila nilai KMO lebih dari 0,5 maka asumsi sampel mewakili populasi terpenuhi atau sampel representatif. Uji KMO seperti pada persamaan 3.1 (Mahmudan, 2020).

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (3.1)$$

Dengan:

p = banyaknya variabel

r_{ij} = koefisien korelasi antara variabel i dan j

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

3.2.2 Multikolinieritas

Multikolinieritas adalah suatu peristiwa dimana terjadi korelasi yang kuat antara dua atau lebih variabel kelompok. Multikolinieritas merupakan masalah yang perlu diperhatikan dalam analisis multivariat pada umumnya, karena pengaruhnya yang sangat besar dalam menghasilkan solusi, sehingga mengganggu proses analisis (Hair, 1998). Untuk mengetahui adanya multikolinieritas salah satunya adalah dengan melihat nilai korelasi antar variabel independen dengan persamaan 3.2 (Gujarati, 2006). Multikolinieritas terjadi jika koefisien korelasi antar variabel lebih dari 0.85 (Widarjono, 2009).

$$r = \frac{n \sum(xy) - (\sum x \sum y)}{(n \sum x^2 - \sum x)^2 ((n \sum y^2 - \sum y)^2)} \quad (3.2)$$

dengan:

r = nilai koefisien korelasi

x = nilai variabel pertama

y = nilai variabel kedua

n = jumlah data

3.3. *Principal Component Analysis (PCA)*

PCA merupakan kombinasi linear dari variabel awal yang secara geometris, kombinasi linear ini merupakan sistem koordinat baru yang diperoleh dari rotasi sistem semula. Metode PCA sangat berguna digunakan jika data yang dimiliki

jumlah variabelnya besar dan memiliki korelasi antar variabelnya. Perhitungan dari *principal component analysis* didasarkan pada perhitungan nilai *eigen* dan *eigenvector* yang penyebaran data dari suatu dataset (Johnson & Wichen, 2007). Tujuan dari analisis PCA adalah untuk mereduksi variabel yang ada menjadi lebih sedikit tanpa harus kehilangan informasi yang termuat dari data asli.

Langkah-langkah untuk melakukan analisis PCA yaitu pertama dengan melakukan standarisasi data. Langkah ini bertujuan agar interval data menjadi lebih proporsional dengan menyamakan rentang nilai pada data. Berikut merupakan persamaan untuk melakukan standarisasi data dengan *Z-Score*.

$$Z = \frac{x - \mu}{\sigma} \quad (3.3)$$

dengan:

Z : nilai *standard score*

x : data observasi

μ : rata-rata variabel

σ : standar deviasi

Kemudian menghitung matriks varian kovarian sebagai langkah awal dalam melakukan analisis PCA untuk mendapatkan kedekatan hubungan antar variabel penelitian. Selanjutnya, menghitung *eigenvalue* dan *eigenvector* yang didapatkan dari matriks kovarian yang diperoleh. Perhitungan nilai eigen dengan menggunakan persamaan berikut (Muhtadi, 2017).

$$[\lambda I - A] = 0 \quad (3.4)$$

dengan λ adalah *eigenvalue*, I adalah matriks identitas, dan A adalah matriks data. Kemudian menghitung eigen vector dengan menggunakan persamaan berikut (Muhtadi, 2017).

$$[\lambda I - A][X] = 0 \quad (3.5)$$

dengan λ adalah *eigenvalue*, I adalah matriks identitas, dan A adalah matriks data, dan X adalah *eigenvector*. Setelah itu, mengurutkan *eigenvalue* secara *descending* dan *Principal Component* (PC) didapatkan dari deretan *eigenvector*. Jumlah *principal component* yang terpilih adalah nilai $\lambda > 1$. Selanjutnya, didapatkanlah dataset baru.

3.4. Penentuan Jumlah *Cluster* Optimum

Terdapat beberapa pendekatan untuk menentukan jumlah *cluster* optimum yaitu terdapat tiga pendekatan diantaranya adalah indeks *connectivity*, indeks *dunn*, dan *silhouette*. Formula dari masing-masing pendekatan indeks tersebut sebagai berikut (Irwansyah & Faisal, 2015).

a. Indeks *Connectivity*

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L X_{i,nni(j)} \quad (3.6)$$

Dimana $nni(j)$ merupakan pengamatan terdekat i ke j dan L merupakan parameter yang menentukan jumlah tetangga yang berkontribusi pada pengukuran *connectivity*.

b. Indeks *Dunn*

Indeks *dunn* adalah rasio jarak terkecil antara observasi pada *cluster* yang berbeda dengan jarak terbesar pada masing-masing *cluster* data.

$$D = \frac{\min_{1 \leq i < j \leq n} d(i, j)}{\max_{1 \leq k \leq n} d'(k)} \quad (3.7)$$

Dimana i, j , dan k merupakan masing-masing indeks untuk *cluster*, d mengukur jarak antar *cluster*, dan d' mengukur perbedaan masing-masing *cluster*.

c. Indeks *Silhouette*

Indeks *Silhouette* adalah dengan mengukur derajat kepercayaan dalam proses *clustering* pada pengamatan baru, jika dikatakan baik dengan nilai indeks mendekati 1 dan kondisi sebaiknya jika nilai indeks mendekati -1.

$$s(i) = \frac{b(i) - a(i)}{\max(a_i, b_i)} \quad (3.8)$$

Dimana $a(i)$ merupakan jarak rata-rata antara i dengan data lainnya pada *cluster* yang sama, dan $b(i)$ merupakan jarak rata-rata antara i dengan data lainnya pada *cluster cluster* yang berbeda.

3.5. Analisis *Cluster*

Analisis *cluster* merupakan analisis teknik multivariat yang bertujuan untuk melakukan kelompok penempatan sekumpulan objek dalam dua atau lebih kelompok berdasarkan kesamaan objek berdasarkan berbagai karakteristik (Simamora, 2005). Analisis *cluster* mengelompokkan elemen mirip sebagai objek

penelitian yang mempunyai tingkat homogenitas yang tinggi antar objek menjadi *cluster* yang berbeda dengan tingkat heterogenitas obyek yang tinggi antar *cluster*.

3.6. Non Hierarchial Methods (Metode Non Hierarchical)

Metode *non hierarchical* merupakan pengelompokkan objek dimana banyaknya *cluster* yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur pengerombolan. Pengelompokan berbasis *partitioning* menghasilkan sebuah partisi dari data mengakibatkan objek dalam kelompok lebih mirip satu sama lain daripada objek yang ada dalam kelompok lain (Triaynto, 2015). Terdapat beberapa metode yang termasuk dalam *Clustering Non Hierarchical* salah satunya yaitu *Clustering K-Medoids*.

3.6.1 Clustering K-Medoids

Algoritma K-Medoids atau sering disebut juga dengan algoritma PAM (*Partitioning Arround Medoid*) dikembangkan oleh Leonard Kaufman dan Peter J.Rousseeuw yang dimana algoritma ini mirip dengan *k-means*, terutama keduanya merupakan algoritma partitional dengan kata lain, keduanya memecah dataset menjadi kelompok-kelompok. Perbedaan dari algoritma *k-means* dengan *k-medoids* adalah terletak pada penentuan pusat *cluster*, yang dimana algoritma *k-medoids* menggunakan objek data perwakilan (*medoids*) sebagai pusat *cluster* (Kaur, Kaur, & Singh). Algoritma *k-medoids* memiliki kelebihan untuk mengatasi kelemahan pada *k-means* yang sensitif terhadap *outlier*, yang dimana objek dengan nilai yang besar memungkinkan menyimpang pada dari distribusi data.

K-Medoids Clustering menggunakan metode pengclusteran partisi untuk mengclusterkan sekumpulan n objek menjadi sejumlah k *cluster*. Algoritma ini menggunakan objek pada kumpulan objek yang mewakili sebuah *cluster*. Objek yang mewakili sebuah *cluster* disebut dengan *medoids*. *Cluster* dibangun dengan menghitung kedekatan yang dimiliki antara *medoids* dengan objek *non medoids* (Setyawati, 2017). Tahapan *K-Medoids Clustering* adalah sebagai berikut:

1. Secara acak pilih k objek pada sekumpulan n objek sebagai *medoids*
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidean Distance* dengan persamaan:

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; 1, 2, 3, \dots, n \quad (3.9)$$

dengan:

x_i = nilai objek i pada variabel x

y_i = nilai objek i pada variabel y

3. Pilih secara acak objek masing-masing *cluster* sebagai i kandidat medoid baru.
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoid baru
5. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru-total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai medoid
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

3.7. Hierarchical Methods (Metode Hierarchical)

Metode *hierarchical* digunakan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifatnya dan *cluster* yang diinginkan belum diketahui (Matjik & Sumertajaya, 2011). Metode *hierarchical* pengelompokkannya dimulai dengan dua objek atau lebih yang memiliki kesamaan terdekat, yang kemudian perhitungan akan diteruskan ke objek lainnya yang memiliki kedekatan selanjutnya atau kedua. Begitupun seterusnya hingga membentuk dendrogram dengan terdapat *hierarchical* (tingkatan) yang jelas antar objeknya. Hasil pengclusteran dengan metode *hierarchical* dapat digunakan dalam sebuah diagram pohon yang disebut dengan dendrogram. Dendrogram digunakan untuk membantu memperjelas proses *hierarchical* tersebut (Hair, Black, J.Babin, & Anderson, 2009). Terdapat dua teknik *hierarchical*, yaitu teknik *agglomerative* (teknik penggabungan) dan *divise* (teknik pembagian). Terdapat beberapa jenis metode Hierarki sebagai berikut:

- 1) *Single Linkage* merupakan jarak dua *cluster* yang diukur dengan jarak terdekat antara sebuah objek dalam *cluster* yang satu dengan sebuah objek dalam *cluster* yang lain.

$$d(uv)w = \min (d_{uw}d_{vw}) \quad (3.10)$$

dengan:

$d(uv)w$ = jarak antara *cluster* (UV) dan *cluster* W

d_{uw} dan d_{vw} = jarak antara tetangga terdekat *cluster* U dan W, serta *cluster* V dan W

- 2) *Complete linkage* merupakan jarak dua *cluster* yang diukur dengan jarak terjauh antara sebuah objek dalam *cluster* yang satu dengan sebuah objek dalam *cluster* yang lain.

$$d(uv)w = \max (d_{uw}d_{vw}) \quad (3.11)$$

- 3) *Average linkage* merupakan jarak antara dua *cluster* yang diukur dengan jarak rata-rata antara sebuah objek dalam *cluster* yang satu dengan sebuah objek dalam *cluster* yang lain.

$$d(uv)w = \frac{\sum_i \sum_k d_{ik}}{N_{(uv)}N_w} \quad (3.12)$$

dengan:

d_{ik} = jarak antara objek ke-i dalam *cluster* (UV) dan objek ke-k dalam *cluster* W

$N_{(uv)}$ dan N_w = jumlah objek dalam *cluster* (UV) dan W

- 4) Metode *Ward* adalah metode pengclusteran bersifat agglomerative untuk memperoleh kelompok yang memiliki varian internal sekecil mungkin. Untuk penclusteran metode *ward*, ukuran yang digunakan adalah *Sum of Square* (SSE).

$$SSE = \sum_{i=1}^n (X_i - \bar{x})'(X_i - \bar{x}) \quad (3.13)$$

dengan:

d_{ik} = jarak antara objek ke-i dalam *cluster* (UV) dan objek ke-k dalam *cluster* W

$N_{(uv)}$ dan N_w = jumlah objek dalam *cluster* (UV) dan W

BAB IV

METODOLOGI PENELITIAN

4.1. Populasi Penelitian

Populasi dalam penelitian ini merupakan data jumlah tindakan kriminalitas dengan jumlah populasi 34 Provinsi di Indonesia. Sampel yang digunakan adalah Data Tindakan Kriminalitas pada Tahun 2020.

4.2. Metode Pengambilan Data

Sumber data yang digunakan dalam penelitian berasal dari *website* resmi Badan Pusat Statistika (BPS). Sedangkan jenis data yang digunakan dalam penelitian adalah data sekunder berasal dari Badan Pusat Statistika yaitu Data Kriminalitas Tahun 2020 yang bersumber dari Biro Pengendalian Operasi Mabes Polri.

4.3. Variabel penelitian

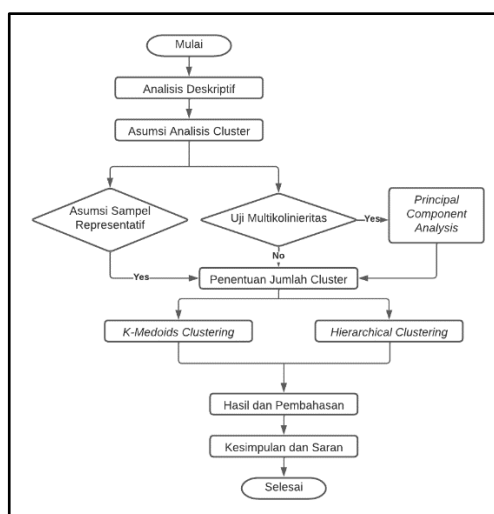
Tabel 4.1 Variabel Penelitian

No	Variabel	Kode	Definisi
1	Kejahatan Terhadap Nyawa	X1	Klasifikasi kejahatan ini meliputi pembunuhan
2	Kejahatan Terhadap Fisik	X2	Klasifikasi kejahatan ini meliputi penganiyaan berat, penganiyaan ringan, dan kekerasan dalam rumah tangga
3	Kejahatan Terhadap Kesusilaan	X3	Klasifikasi kejahatan ini perkosaan dan pencabulan
4	Kejahatan Terhadap Kemerdekaan Orang	X4	Klasifikasi kejahatan ini meliputi penculikan dan mempekerjakan anak dibawah umur
5	Kejahatan Terhadap Hak Milik dengan Penggunaan Kekerasan	X5	Klasifikasi kejahatan ini meliputi Pencurian dengan Kekerasan, Pencurian dengan Kekerasan Menggunakan Senjata Api (Senpi), Pencurian dengan Kekerasan Menggunakan Senjata Tajam (Sajam)

6	Kejahatan Terhadap Hak Milik dengan Tanpa Penggunaan Kekerasan	X6	Kejahatan ini meliputi pencurian, pencurian kendaraan bermotor, pencurian dengan pemberatan, pembakaran dengan senjata
7	Kejahatan Terkait Narkotika	X7	Kejahatan ini terkait narkotika dan psikotropika
8	Kejahatan Terkait Penipuan, Penggelapan, Korupsi	X8	Kejahatan ini meliputi penipuan, penggelapan uang, dan korupsi

4.4. Metode Analisis Data

Dalam penelitian ini menggunakan alat bantu yang digunakan untuk mencapai tujuan dalam analisis ini yaitu perangkat lunak *Software R-Studio*, *Microsoft Excel*. Tahapan *flowchart* nya yaitu Analisis deskriptif yang bertujuan untuk memberikan gambaran untuk mengenai data yang digunakan. Kemudian tahap selanjutnya Uji multikolinieritas, untuk melihat terjadi korelasi yang kuat antara dua atau variabel kelompok, jika terdapat multikolinieritas maka dilanjutkan dengan analisis *Principal Component Analysis* (PCA). Selanjutnya, menentukan jumlah *cluster* yaitu menggunakan *K-Medoids Clustering* dan *Hierarchical Clustering*, kemudian hasil dan pembahasan, tahap terakhir yaitu kesimpulan dan saran.



Gambar 4.1 Flowchart

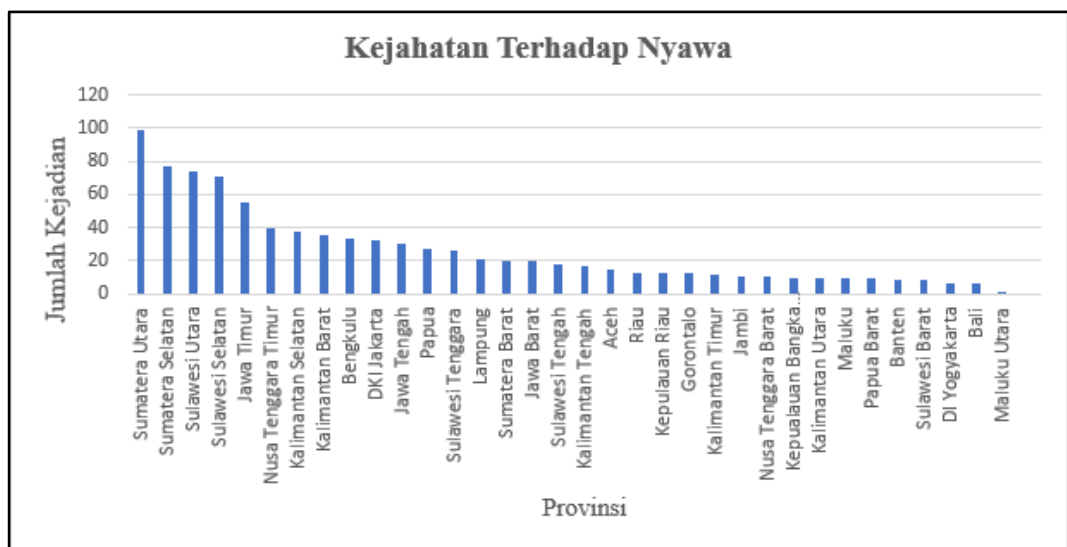
BAB V

HASIL DAN PEMBAHASAN

5.1. Analisis Deskriptif

Tujuan dari analisis deskriptif adalah untuk mempermudah pembaca dalam memahami suatu gambaran dari sebuah data. Pada penelitian ini dilakukan analisis deskriptif mengenai data kriminalitas menurut kepolisian provinsi di Indonesia tahun 2020. Grafik dari masing-masing variabel sebagai berikut.

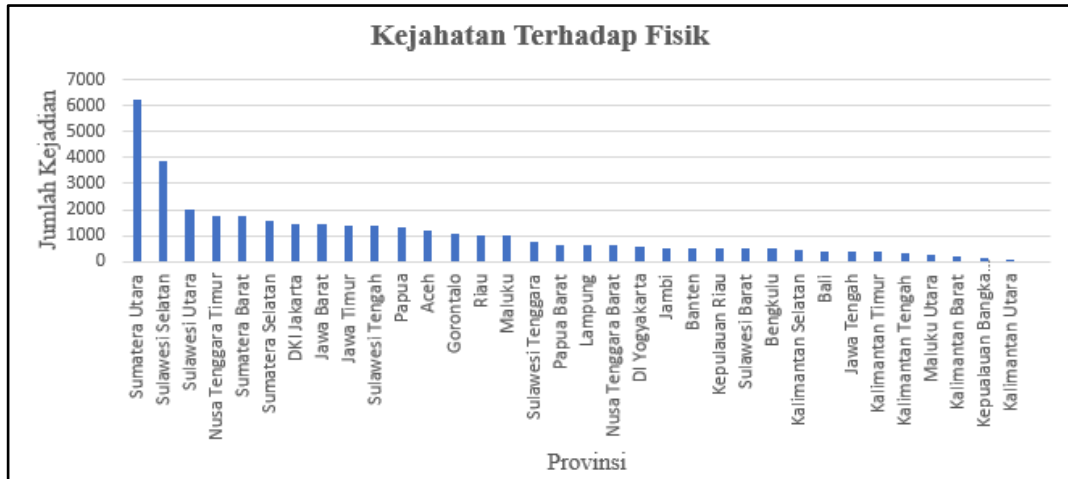
a. Kejahatan Terhadap Nyawa



Gambar 5.1 Kejahatan Terhadap Nyawa

Gambar 5.1 dapat dilihat bahwa terdapat tiga poldata kepolisian di masing-masing provinsi dengan kasus terbanyak yaitu Sumatera Utara (99 kejadian), Sumatera Selatan (77 Kejadian), dan Sulawesi Utara (74 Kejadian). Selain itu juga terdapat wilayah yang sedikit yaitu Maluku Utara dengan jumlah kejadian sebanyak satu kejadian. Fenomena yang menyebabkan angka kejahatan terhadap nyawa naik yaitu dikarenakan terjadi pandemi covid-19 yang diharuskan masyarakat melakukan sosial *distancing* dan dilakukannya sistem Pemberlakuan kebijakan pembatasan sosial berskala besar (PSBB) yang membuat masyarakat menjadi memiliki tekanan jiwa karena kurangnya interaksi antar individu.

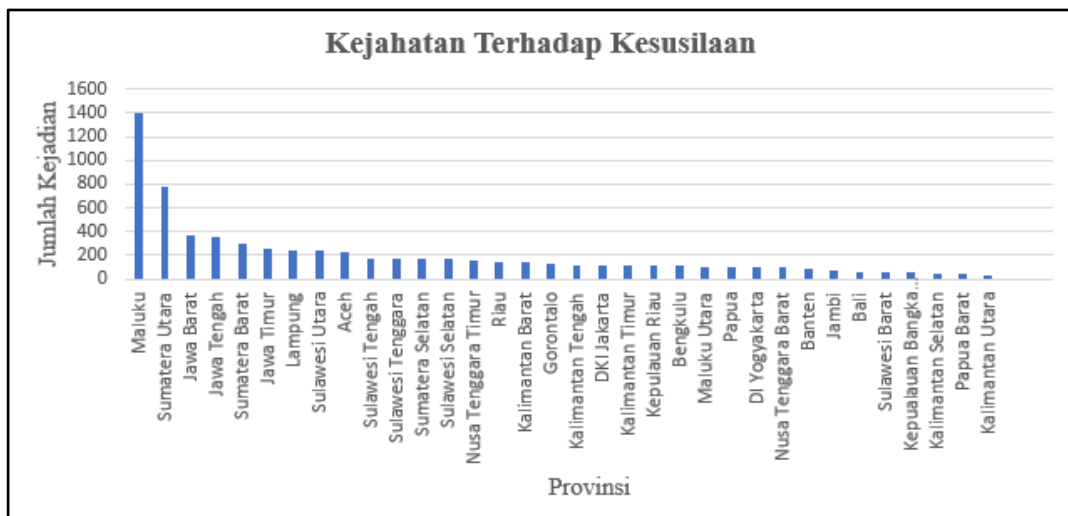
b. Kejahatan Terhadap Fisik



Gambar 5.2 Kejahatan Terhadap Fisik

Gambar 5.2 menunjukkan kejadian terhadap fisik, dimana kejadian yang terbanyak terjadi pada polda kepolisian Sumatera Utara sebanyak 6207 kejadian, diikuti dengan kejadian terbanyak kedua yaitu pada Sulawesi Selatan sebanyak 3848 kejadian, dan ketiga terbanyak yaitu pada Sulawesi Utara sebanyak 1976 kejadian. Selain itu juga terdapat polda kepolisian dengan kejadian terhadap fisik paling sedikit yaitu Kepulauan Bangka Belitung dan Kalimantan Utara masing-masing sebanyak 150 dan 91 kejadian. Faktor penyebab terjadinya kejahatan terhadap fisik yaitu dengan meningkatnya kasus covid-19 yang menyebabkan tingginya gelombang PHK menyebabkan hilangnya mata pencaharian yang berdampak pada meningkatnya beban keluarga dan stress yang berpotensi memicu kekerasan dalam rumah tangga (KDRT) dan kekerasan berbasis gender lainnya.

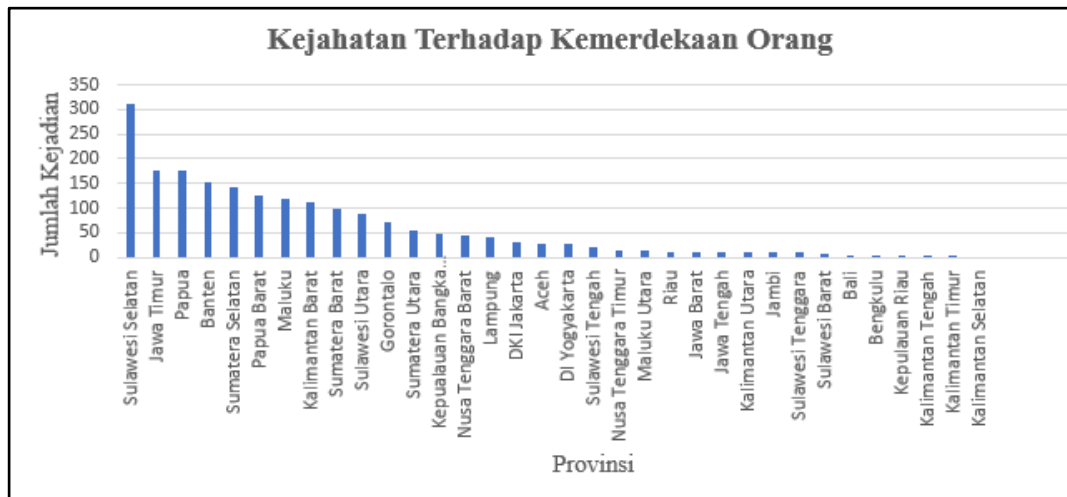
c. Kejahatan Terhadap Kesusilaan



Gambar 5.3 Kejahatan Terhadap Kesusilaan

Dari gambar 5.3 menunjukkan kejahatan terhadap kesusilaan yang dimana pada tahun 2020 kejadian terbanyak di Indonesia terjadi pada polda kepolisian Maluku dengan kejadian sebanyak 1398 kejadian. Selain itu juga wilayah pada urutan kedua dan ketiga diikuti oleh wilayah Sumatera Utara dan Jawa Barat yang mana masing-masing kejadian sebanyak 774 kejadian dan 371 kejadian. Wilayah dengan jumlah polda kepolisian paling sedikit yaitu Polda Kalimantan Selatan, Papua Barat dan Kalimantan Utara dengan masing-masing kejadian sebanyak 51 kejadian, 47 kejadian, dan 32 kejadian. Fenomena yang terjadi yaitu pemerkosaan, yang mana pada tahun 2020 terjadi pandemi dan menyebabkan kasus kejahatan kesusilaan meningkat yang mana para predator mengincar anak-anak dan perempuan dari dampak penggunaan internet dan sosial media yang tidak bijak.

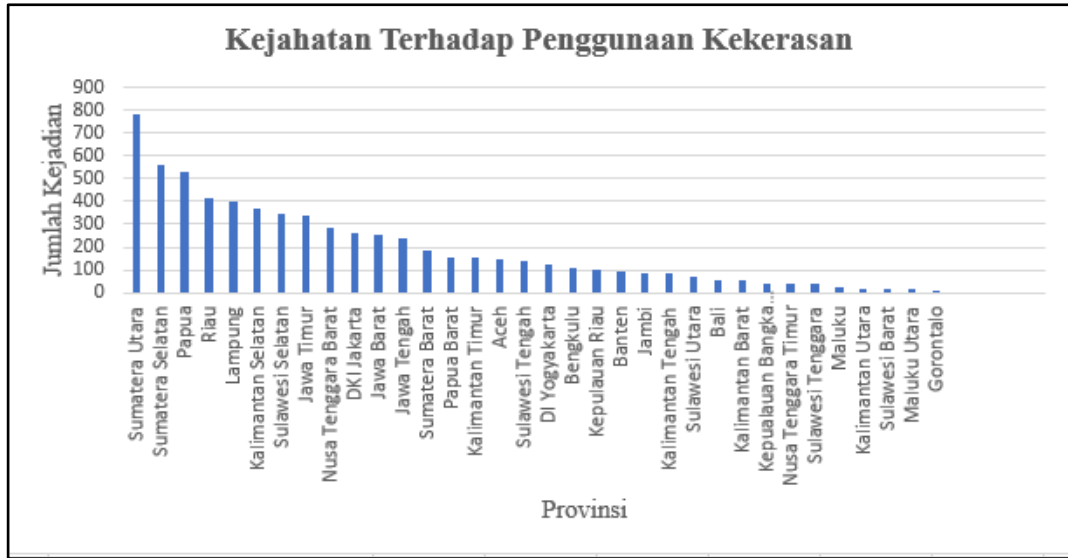
d. Kejahatan Terhadap Kemerdekaan Orang



Gambar 5.4 Kejahatan Terhadap Kemerdekaan Orang

Dari gambar 5.4 menunjukkan kejadian terhadap kemerdekaan orang tahun 2020 berdasarkan polda/provinsi. Diketahui bahwa polda Sulawesi Selatan merupakan kejadian terbanyak yaitu dengan kejadian sebanyak 311. Selain itu di urutan kedua terdapat polda Jawa Timur sebanyak 177 kejadian, dan urutan ketiga terdapat polda Papua sebanyak 176 kejadian. Sementara itu, Polda Kalimantan Selatan termasuk kejadian paling sedikit karena tidak terdapat kejahatan ini di tahun 2020. Salah satu faktor penyebab mempekerjakan anak dibawah umur adalah karena pembelajaran jarak jauh (PJJ) yang dilakukan selama covid-19 juga beresiko. Hal tersebut karena telah memunculkan adanya ketimpangan akses teknologi informasi disamping krisis lainnya.

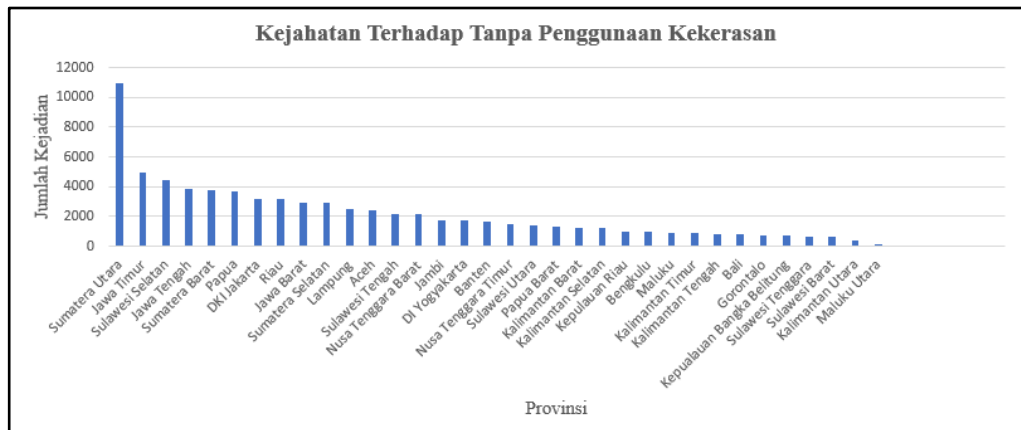
e. Kejahatan Terhadap Penggunaan Kekerasan



Gambar 5.5 Kejahatan Terhadap Penggunaan Kekerasan

Gambar 5.5 menunjukkan bahwa jumlah kejadian kejahatan terhadap hak/milik barang dengan penggunaan kekerasan di tahun 2020. Wilayah dengan kejadian terbanyak yaitu terjadi pada polda Sumatera Utara sebanyak 780 kejadian, urutan kedua terbanyak adalah polda Sumatera Selatan sebanyak 563 kejadian. Selain itu, dua wilayah dengan kejadian paling sedikit terjadi di Polda Maluku Utara dan Gorontalo dengan masing-masing kejadian sebanyak 14 dan 4 kejadian. Fenomena yang terjadi yaitu salah satunya kasus pencurian dengan bersenjata tajam yang mana pada tahun 2020 masa pandemi yang menyebabkan masyarakat dalam keadaan ekonomi yang sulit, maka dari itu orang jahat memiliki kesempatan untuk memenuhi kebutuhannya dengan mencuri, menjambret, dll.

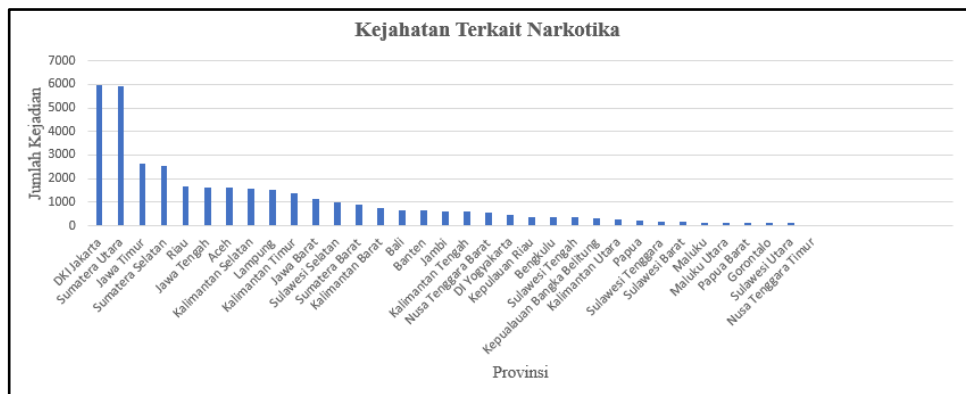
f. Kejahatan Terhadap Tanpa Penggunaan Kekerasan



Gambar 5.6 Kejahatan Terhadap Tanpa Penggunaan Kekerasan

Gambar 5.6 adalah kejadian kejahatan hak/milik tanpa penggunaan kekerasan yang mana pada urutan pertama terjadi pada Polda Sumatera Utara yaitu sebanyak 10916 kejadian, urutan kedua terjadi Polda Jawa Timur sebanyak 4976 kejadian. Selain itu juga terdapat wilayah dengan kejadian paling sedikit yaitu pada Polda Kalimantan Utara dan Maluku Utara dengan masing-masing kejadian sebanyak 340 dan 112 kejadian. Fenomena pada kejahatan ini dikarenakan pandemi yang menyebabkan tingkat pengangguran yang tinggi dan kinerja pengamanan yang kurang memadai.

g. Kejadian Terkait Narkotika



Gambar 5.7 Kejahatan Terkait Narkotika

Gambar 5.7 menunjukkan bahwa kejadian kejahatan terkait narkotika di tahun 2020. Pada kejadian ini untuk paling banyak pertama terdapat di wilayah DKI Jakarta (Polda Metro Jaya) yaitu sebanyak 5981 kejadian, urutan kedua adalah Sumatera Utara sebanyak 5932 kejadian, dan urutan ketiga yaitu Jawa Timur dengan jumlah sebanyak 2629 kejadian. Sementara itu, terdapat juga kejadian paling sedikit yaitu tiga terendah terjadi pada Polda Gorontalo, Sulawesi Utara, dan Nusa Tenggara Timur dengan jumlah kejadian masing-masing sebanyak 101 kejadian, 100 kejadian, dan 41 kejadian. Penyebab terjadinya narkotika adalah faktor himpitan ekonomi pada masa pandemi karena adanya PHK ataupun tidak memiliki pekerjaan tetap maupun adanya pembatasan kegiatan masyarakat yang menyebabkan pelaku nekad mengedarkan narkotika untuk memenuhi kebutuhannya, selain itu juga faktor lingkungan sosial.

h. Kejahatan Terhadap Penipuan, Penggelapan, Korupsi



Gambar 5.8 Kejahatan Terhadap Penipuan, Penggelapan, Korupsi

Gambar 5.8 menunjukkan kejadian kehajatan terkait penipuan, penggelapan, dan korupsi terbanyak yaitu pada Polda Sumatera Utara di urutan pertama dengan jumlah kejadian sebanyak 5562 kejadian, urutan kedua di wilayah Polda DKI Jakarta (Metro Jaya) yaitu sebanyak 4595 kejadian. Gambar 6.8 juga menunjukkan kejadian paling sedikit dengan dua wilayah diantaranya Polda Kalimantan Utara dan Maluku Utara dengan masing-masing jumlah kejadian sebanyak 72 dan 49 kejadian. Faktor penyebab terjadinya korupsi karena pandemi covid-19 meningkat yang membuat masyarakat menjadi serakah dan mengambil kesempatan dalam pengumpulan dana bansos dengan mengambil uang orang lain yang bukan miliknya.

i. Tingkat kejahatan masing-masing provinsi

Dari banyaknya jenis kejahatan yang ada di Indonesia terdapat beberapa provinsi yang menonjol dikarenakan jumlah banyaknya kasus tersebut, yaitu Kepolisian Sumatera Utara, karena dari masing-masing kejahatan termasuk dalam kasus kejahatan terbanyak yaitu kejahatan terhadap nyawa, kejahatan terhadap fisik, kejahatan terhadap kemerdekaan orang, kejahatan tanpa terhadap penggunaan kekerasan, kejahatan terhadap penggunaan kekerasan, kejahatan terkait narkoba, dan kejahatan terkait penipuan, penggelapan, korupsi. Selain itu yang menonjol lainnya yaitu Jawa Timur jika dilihat dari sisi kejahatan terbanyak yaitu kejahatan terhadap kemerdekaan orang, kejahatan tanpa penggunaan kekerasan, kejahatan terkait narkoba, kejahatan terkait penipuan, penggelapan, korupsi.

5.2. Uji Asumsi Cluster

5.2.1 Uji Kaiser-Meyer Olkin

Uji KMO digunakan untuk mengukur kecukupan pengambilan sampel secara keseluruhan dan mengukur kecukupan pengambilan sampel untuk setiap indikator. Apabila nilai KMO lebih dari 0,5 maka asumsi sampel mewakili populasi terpenuhi atau sampel representatif. Hasil nilai KMO dapat dilihat pada Tabel 5.1.

Tabel 5.1 Koefisien Korelasi

<i>Kaiser Mayer Olkin</i> (KMO)	0.78
------------------------------------	------

Berdasarkan hasil nilai KMO pada tabel 5.1, didapatkan nilai KMO sebesar 0,78 yang artinya nilai ini melebihi 0,5 yang berarti bahwa sampel mewakili populasi atau sampel representatif sehingga analisis bisa dilanjutkan ke tahap selanjutnya.

5.2.2 Uji Multikolinieritas

Uji multikolinieritas merupakan terjadinya korelasi yang kuat antara dua atau lebih variabel kelompok, salah satu masalah yang harus diperhatikan dalam menganalisis multivariat, karena multikolinieritas disebabkan oleh pengaruh besar dalam menghasilkan solusi sehingga nantinya dapat mengganggu proses analisis.

Tabel 5.2 Koefisien Korelasi

	X1	X2	X3	X4	X5	X6	X7	X8
X1	1.00	0.77	0.23	0.43	0.63	0.67	0.53	0.61
X2	0.77	1.00	0.41	0.40	0.63	0.85	0.57	0.71
X3	0.23	0.41	1.00	0.16	0.17	0.36	0.21	0.30
X4	0.43	0.40	0.16	1.00	0.31	0.32	0.03	0.15
X5	0.63	0.63	0.17	0.31	1.00	0.82	0.66	0.67
X6	0.67	0.85	0.36	0.32	0.82	1.00	0.73	0.84
X7	0.53	0.57	0.21	0.38	0.66	0.73	1.00	0.91
X8	0.61	0.71	0.30	0.15	0.67	0.84	0.91	1.00

Pada tabel 5.2 diatas diketahui bahwa terdapat nilai korelasi > 0.85 maka hal itu menandakan terjadinya multikolinieritas. Dari *output* matriks korelasi pada tabel 5.2 terdapat beberapa variabel lain yang mengandung multikolinieritas, namun terdapat pula beberapa variabel yang tidak mengandung multikolinieritas. Variabel yang mengandung multikolinieritas yaitu Kejahatan Terkait Narkotika (X7) dengan Kejahatan Terhadap Penipuan, Penggelapan, Korupsi (X8).

5.3. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) atau analisis komponen utama merupakan solusi jika dalam proses analisis *clustering* terjadi multikolinieritas. Tujuan dari analisis PCA adalah untuk mereduksi variabel yang ada menjadi lebih sedikit tanpa harus kehilangan informasi yang termuat dari data asli. Dalam menentukan jumlah komponen utama yang dihasilkan, dapat dilihat dari *eigenvalue* yang menunjukkan lebih dari satu. Hasil PCA seperti pada Tabel 5.3.

Tabel 5.3 Tabel *Eigenvalues*

Komponen	Total	% of Variance	Cumulative %
1	4,831	60,38	60,38
2	1,131	14,15	74,53
3	0,899	11,24	85,77
4	0,415	5,199	90,97
5	0,344	4,307	95,27
6	0,253	3,163	98,44
7	0,075	0,949	99,39
8	0,048	0,609	100

Dari tabel 5.3 menunjukkan nilai *eigenvalue* dan proporsi kumulatif untuk tiap faktornya. Dari hasil tersebut, komponen 1 dan 2 memiliki nilai *eigenvalue* lebih dari satu. Sehingga dipilihlah komponen yang terbentuk dari analisis PCA ini sebanyak 2 komponen utama. Komponen pertama memiliki *eigenvalue* sebesar 4,830 dengan varian sebesar 60,38%. komponen kedua memiliki nilai *eigenvalue* sebesar 1,131 dengan varian sebesar 14,15%. Dengan menggunakan 2 komponen utama, kedua komponen tersebut telah mampu menjelaskan 74,53% keragaman data. Dibawah ini merupakan hasil dari PCA.

Tabel 5.4 Hasil PCA

PC1	PC2
-0,076	0,525
9,198	1,096
0,544	-0,588
0,427	0,966
-1,178	0,600

2,692	-0,628
-1,120	0,309
0,476	0,538
-1,854	-0,014
-1,418	0,488
2,884	2,687
0,921	0,829
0,647	0,829
-0,985	0,495
2,805	-0,062
-0,880	0,922
-1,692	0,615
-0,519	0,249
-0,595	-0,022
-0,863	-0,804
-1,510	0,482
-0,181	0,808
-1,221	0,792
-2,149	0,355
0,454	-1,324
-0,576	0,091
3,317	-3,258
-1,459	0,007
-1,399	-0,525
-1,932	0,340
-0,372	-2,173
-2,227	0,271
-1,101	-0,871
0,996	-1,421

5.4. Penentuan Jumlah *Cluster* Optimum

Pada penelitian ini menggunakan tiga pendekatan yaitu indeks *connectivity*, indeks *dunn*, dan *silhouette*. Pengambilan jumlah *cluster* dengan melihat nilai indeks *connectivity* paling kecil, nilai indeks *dunn* paling besar, dan nilai *silhouette* mendekati 1. Berikut merupakan hasil yang didapatkan dari validasi *cluster* *k-medoids* dan *hierarchical* seperti pada tabel 5.5.

Tabel 5.5 Penentuan Jumlah *Cluster* Optimum

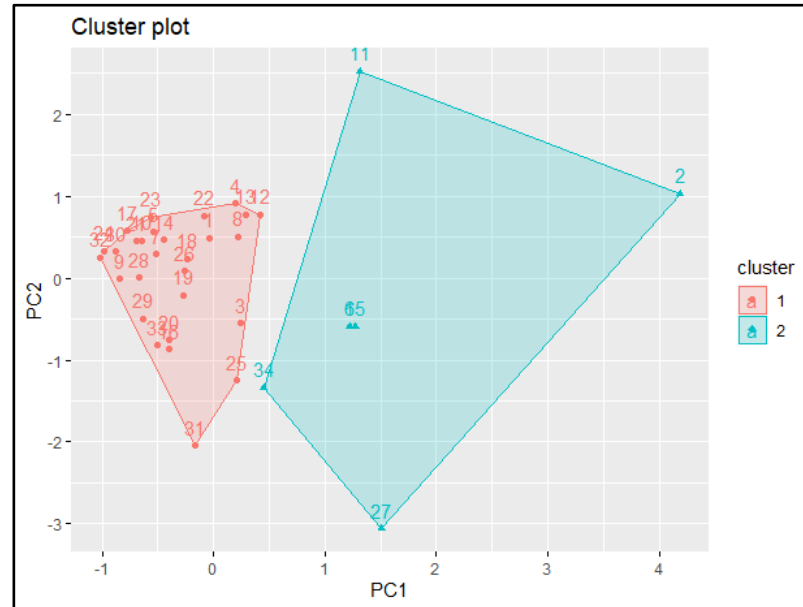
Metode <i>Cluster</i>	Indeks	Jumlah <i>Cluster</i>			
		2	3	4	5
<i>K-Medoids Clustering</i>	<i>Connectivity</i>	11,204	18,467	16,072	26,280
	<i>Dunn</i>	0,064	0,060	0,135	0,078
	<i>silhouette</i>	0,582	0,375	0,375	0,384
<i>Hierarchical Clustering</i>	<i>Connectivity</i>	2,929	5,857	11,287	13,037
	<i>Dunn</i>	0,984	0,409	0,508	0,508
	<i>silhouette</i>	0,735	0,502	0,545	0,506

Berdasarkan pada hasil tiga pendekatan di tabel 5.5, diperoleh bahwa pada algoritma *k-medoids* dilihat dari nilai indeks *connectivity* yang paling kecil dengan nilai 11,204 pada *cluster* 2, nilai *dunn* dengan nilai paling besar yaitu 0,135 pada *cluster* 4, dan nilai *silhouette* yang mendekati 1 yaitu 0,582 pada *cluster* 2. Maka pada algoritma *k-medoids* diperoleh *cluster* yang baik sebanyak 2. Selanjutnya yaitu pada algoritma *hierarchical clustering* dilihat dari nilai indeks *connectivity* yang paling kecil dengan nilai 2,929 pada *cluster* 2, nilai *dunn* dengan nilai paling besar yaitu 0,98 pada *cluster* 2, dan nilai *silhouette* yang mendekati 1 yaitu 0,73 pada *cluster* 2. Sehingga, pada algoritma *hierarchical clustering* diperoleh *cluster* yang paling baik sebanyak 2. Maka pada penelitian ini menggunakan jumlah *cluster* sebanyak 2.

5.5. *K-Medoids Clustering*

K-medoids adalah metode *Non Hierarchical* yang merupakan varian dari dari algoritma *K-Means*. Dalam menentukan sekumpulan *n* objek menjadi sejumlah *k cluster*, *k-medoids* menggunakan metode pengelompokkan partisi. Algoritma ini menggunakan objek pada kumpulan objek yang mewakili sebuah *cluster*. Objek yang mewakili sebuah *cluster* disebut *medoids*. *Cluster* dibangun dengan

menghitung kedekatan yang dimiliki antara objek *medoids* dan *non-medoids*. Dari hasil analisis *cluster* menggunakan algoritma *k-medoids* didapatkan hasil berupa dua *cluster* seperti pada visualisasi berikut:



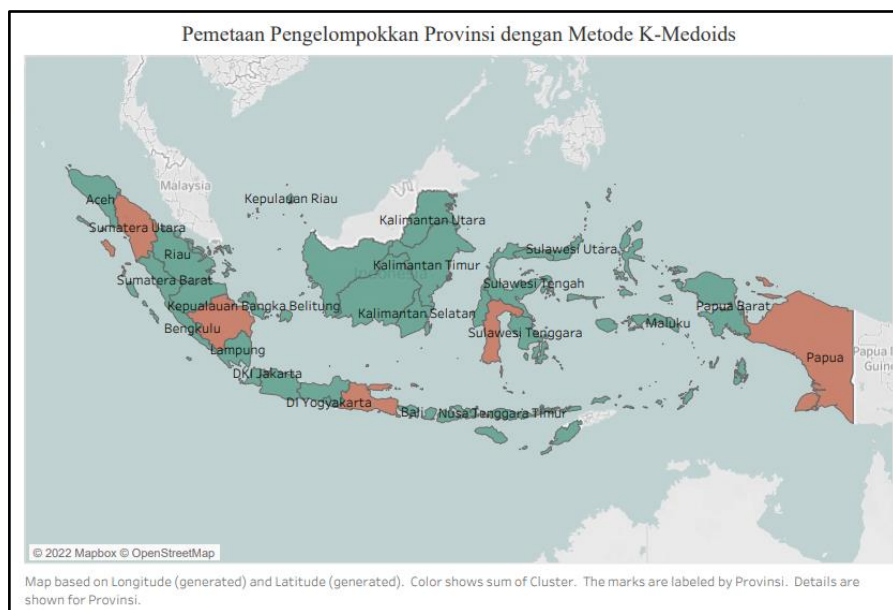
Gambar 5.9 Grafik *K-Medoids Clustering*

Pada gambar 5.9 yaitu hasil plot visualisasi *K-Medoids Clustering* yang terdiri dari dua *cluster* yang dibedakan antara dua warna yaitu merah dan biru. Warna merah yaitu menjelaskan hasil dai *cluster* 1, sedangkan warna biru menjelaskan hasil *cluster* 2. Terlihat bahwa dari masing-masing warna plot terdapat jumlah anggota yang berbeda. Berikut merupakan tampilan hasil *K-Medoids Clustering*.

Tabel 5.6 Hasil *K-Medoids Clustering*

<i>Cluster</i>	Jumlah	Provinsi
1	28	Aceh, Sumatera Barat, Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat
2	6	Sumatera Utara, Sumatera Selatan, DKI Jakarta, Jawa Timur, Sulawesi Selatan, Papua

Dari *cluster* menggunakan algoritma *k-medoids* didapatkan hasil bahwa *cluster* satu terdiri dari 28 kepolisian di Provinsi yaitu Aceh, Sumatera Barat, Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat. Sedangkan *cluster* 2 terdapat 6 kepolisian di Provinsi yang terdiri dari Sumatera Utara, Sumatera Selatan, DKI Jakarta, Jawa Timur, Sulawesi Selatan, Papua .



Gambar 5.10 Pemetaan Kelompok Provinsi

5.6. *Clustering Hierarchical*

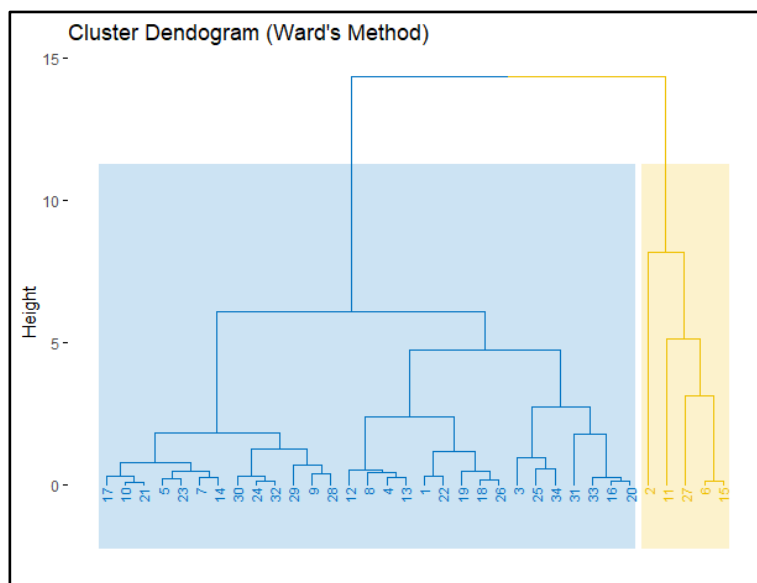
Clustering Hierarchical adalah suatu metode analisis *cluster* yang dilakukan secara bertahap dan bertingkat sehingga membentuk tingkatan seperti pohon. Dalam menentukan jumlah *cluster*, metode ini menghasilkan urutan partisi dengan menggabungkan atau membagi *cluster*. Hasil dari metode ini dapat disajikan dengan dendrogram. Adapun 4 metode *agglomerative* dalam pembagian *cluster* yaitu *single linkage*, *average linkage*, *complete linkage*, dan *ward*. Penentuan metode terbaik pada *cluster hierarchical* dapat ditentukan dengan *agglomerative coefficient*, pada tabel 5.7 merupakan hasil dari *agglomerative coefficient*.

Tabel 5.7 *Agglomerative Coefficient*

Metode	<i>Agglomerative Coefficient</i>
--------	----------------------------------

<i>Single Linkage</i>	0,9051577
<i>Complete Linkage</i>	0,9207571
<i>Average Linkage</i>	0,9175915
<i>Ward</i>	0,9441308

Berdasarkan tabel 5.7 didapatkan bahwa metode *hierarchical* yang menghasilkan *cluster* terbaik yaitu metode *Ward* karena nilai *agglomerative coefficient* lebih besar dibandingkan dengan metode *hierarchical* lainnya. Dendrogram dengan metode *ward* seperti gambar 5.11.



Gambar 5.11 Dendrogram *Ward*

Dendrogram yang diperoleh dari metode *ward* terlihat terpartisi dengan sangat baik. Pada setiap *cluster*, terlihat nilai *height* yang rendah, hal ini dikarenakan cara kerja metode *ward* yaitu meminimumkan nilai *within sum of squared* (*wss*) tiap *cluster*. Pada dendrogram diatas terdapat 29 provinsi pada *cluster* 1, dan terdapat 5 provinsi pada *cluster* 2.

Untuk mengetahui anggota kelompok yang terbentuk pada masing-masing *cluster* dari metode *ward*, seperti pada tabel berikut.

Tabel 5.8 Hasil Pengelompokkan *Ward*

<i>Cluster</i>	Jumlah	Provinsi
1	29	Aceh, Sumatera Barat, Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara

		Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua
2	5	Sumatera Utara, Sumatera Selatan, DKI Jakarta, Jawa Timur, Sulawesi Selatan.

Dari *cluster* menggunakan algoritma Hierarki *clustering* dengan metode *ward* didapatkan hasil bahwa *cluster* 1 terdiri dari 29 kepolisian di Provinsi yaitu Aceh, Sumatera Barat, Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Sedangkan *cluster* 2 terdapat 5 kepolisian di Provinsi yang terdiri dari Sumatera Utara, Sumatera Selatan, DKI Jakarta, Jawa Timur, Sulawesi Selatan.



Gambar 5.12 Pemetaan Pengelompokan Provinsi

5.7. Pemilihan Metode Terbaik

Setelah melakukan proses analisis *clustering*, dilanjutkan dengan melihat metode terbaik menggunakan algoritma *K-Medoids* dan *Ward*. Penelitian ini

menggunakan 3 indeks diantaranya *connectivity*, *dunn*, *silhouette*. Hasil dari ketiga indeks tersebut seperti pada Tabel 5.9.

Tabel 5.9 Hasil Perbandingan Metode Terbaik

Indeks	<i>K-Medoids</i>	<i>Ward</i>
<i>Connectivity</i>	11,204	2,929
<i>Dunn</i>	0,064	0,984
<i>Silhouette</i>	0,582	0,735

Berdasarkan nilai dari masing-masing indeks pada Tabel 5.9 dengan jumlah *cluster* sebanyak 2 menunjukkan bahwa algoritma *Hierarchical Clustering* pada metode *ward* menghasilkan analisis *cluster* terbaik dibandingkan dengan *K-Medoids Clustering*.

5.8. Interpretasi Metode Terbaik

Berdasarkan hasil perbandingan metode *cluster* terbaik menggunakan pendekatan *cluster*, didapatkan hasil yaitu *hierarchial cluster* menggunakan metode *ward* pada data kasus tindakan kriminalitas di Indonesia tahun 2020, berikut merupakan hasil profilasi.

Tabel 5.10 Profilasi *Cluster* Terbaik

<i>Cluster</i>	X1	X2	X3	X4	X5	X6	X7	X8
1	18.89	747	187.5	38.54	132.7	1543	652.5	709.6
2	60.17	2626	264.5	148.5	470.5	5010	3058	2871

Berdasarkan hasil tabel 5.10 didapatkan hasil profilasi atau nilai rata-rata yang artinya dari masing-masing *cluster* memiliki karakteristik yang berbeda. Pada *cluster* 1 dan *cluster* terdapat perbedaan yang mencolok yaitu nilai rata-rata *cluster* 1 lebih rendah dibandingkan *cluster* 2. Warna hijau menunjukkan rendah, dan warna merah menunjukkan tinggi.

Didapatkan hasil provinsi yang termasuk *cluster* 1 yaitu Aceh, Sumatera Barat, Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua termasuk kategori rendah atau dapat dikatakan aman.

Provinsi yang termasuk pada *cluster* 2 yaitu Sumatera Utara, Sumatera Selatan, DKI Jakarta, Jawa Timur, Sulawesi Selatan dengan kasus kriminalitas dengan nilai *cluster* setiap variabel tinggi maka dapat dikatakan provinsi pada *cluster* 2 ini dikategorikan rawan. Dapat dilihat pada situs Sindonews.com, provinsi yang termasuk pada *cluster* 2 termasuk 5 provinsi tingkat kriminalitas tertinggi di Indonesia tahun 2020 (Faizi, 2022).

BAB VI

PENUTUP

6.1. Kesimpulan

1. Dari banyaknya jenis kejahatan yang ada di Indonesia terdapat beberapa provinsi yang menonjol dikarenakan jumlah banyaknya kasus tersebut, yaitu Kepolisian Sumatera Utara, karena dari masing-masing kejahatan termasuk dalam 3 terbanyak yaitu kejahatan terhadap nyawa, kejahatan terhadap fisik, kejahatan terhadap kemerdekaan orang, kejahatan tanpa terhadap penggunaan kekerasan, kejahatan terhadap penggunaan kekerasan, kejahatan terkait narkoba, dan kejahatan terkait penipuan, penggelapan, korupsi.
2. Berdasarkan hasil *cluster* menggunakan algoritma *k-medoids* pada data kriminalitas di Indonesia tahun 2020 didapatkan terdapat 2 *cluster* yaitu diantaranya, *cluster* 1 terdiri dari 28 provinsi, dan *cluster* 2 terdiri dari 6 provinsi di Indonesia.
3. Berdasarkan hasil *cluster* menggunakan algoritma *hierarchial* pada data kriminalitas di Indonesia tahun 2020 didapatkan terdapat 2 *cluster* yaitu diantaranya, *cluster* 1 terdiri dari 29 provinsi, dan *cluster* 2 terdiri dari 5 provinsi di Indonesia.
4. Berdasarkan hasil perbandingan metode *cluster* terbaik didapatkan yaitu menggunakan algoritma *Hierarchial Clustering* metode *Ward* pada data Kriminalitas tahun 2020.

6.2. Saran

Berdasarkan penelitian dan kesimpulan, maka saran yang dapat diberikan sebagai berikut:

1. Untuk provinsi di *cluster* 1 merupakan kasus kejahatan aman, walaupun aman polda harus melakukan tindakan tegas dari aparat hukum agar pelaku menjadi jera.
2. Untuk provinsi di *cluster* 2 dikategorikan sebagai rawan karena jumlah dari semua jenis kejahatan termasuk dalam kategori tinggi. Saran yang harus dilakukan adalah menggerakkan roda perekonomian, dengan tujuan

agar semua orang memiliki kesempatan yang sama untuk berusaha, selain itu jangan mudah terpancing emosi, melakukan peningkatan pendidikan seperti diadakan acara kebudayaan untuk terciptanya sosialisasi

3. Saran untuk penelitian selanjutnya yaitu dengan menggunakan data yang terbaru, menggunakan metode yang lain, dan dapat juga ditambah variabel-variabelnya.

DAFTAR PUSTAKA

- Arby, I. A. (2020). *Daftar 18 Artis yang Terjerat Narkoba dan Psikotropika Sepanjang 2020*. Diambil kembali dari Kompas.com: <https://megapolitan.kompas.com/read/2020/12/05/17113441/daftar-18-artis-yang-terjerat-narkoba-dan-psikotropika-sepanjang-2020?page=all>
- Asia, S. E. (2020). *South-Eastern Asia: Crime Index by Country 2020*. Diambil kembali dari Numbeo: https://www.numbeo.com/crime/rankings_by_country.jsp?title=2020®ion=035
- Dewi, S. M., Windarto, A. P., Damanik, I. S., & H. S. (2019). Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah. *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*.
- Ediyanto, Muhlahsal, & Neva. (2013). PENGKLASIFIKASIAN KARAKTERISTIK DENGAN METODE K-MEANS CLUSTER ANALYSIS. *Buletin Ilmiah Matematika, Statistika dan Terapannya*.
- Faizi, L. (2022, Februari). *5 Provinsi dengan Tingkat Kriminalitas Tertinggi di Indonesia*. Diambil kembali dari SINDONEWS.com: <https://nasional.sindonews.com/read/694639/13/5-provinsi-dengan-tingkat-kriminalitas-tertinggi-di-indonesia-1645610543>
- Gujarati, D. (2006). *Ekonometrika Dasar*. Jakarta: Erlangga.
- Hair. (1998). *Multivariate Data Analysis (Fifth Edition)*. Prentice Hall.
- Hasan, M. I. (2002). *Pokok-Pokok Materi Statistik 2 (Statistik Inferensif)*. Jakarta: Bumi Aksara.
- Irwansyah, E., & Faisal, M. (2015). *ADVANCED CLUSTERING Teori dan Aplikasi*.
- Jain, A. K., & Dubes, R. C. (1998). *Algorithm for Clustering Data*. Englewood Cliffs, New Jersey: Prentice Hall.
- Johnson, & Wichen. (2007). *Applied Multivariate Statistical Analysis Edisi Keenam*. Pearson Pretice Hall.

- Kaur, N. K., Kaur, U., & Singh, D. (t.thn.). K-Medoid *Clustering* Algorithm- A Review. *International Journal of Computer Application and Technology*, 43.
- Lestari, A. D., & dkk. (t.thn.). PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN KRIMINALITAS MENGGUNAKAN METODE WARD DAN K-MEDOIDS. *unimus.ac.id*.
- Mahmudan, A. (2020). *Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering*. *Jurnal Matematika, Statistika, & Komputasi*.
- Mardianto, M. F., & dkk. (2015). Pengelompokan Optimal Kabupaten dan Kota Rawan Kriminalitas di Jawa Timur dengan Metode Analisis Kluster Terbaik. *Zeta - Math Journal Volume 1 No 1*.
- Matjik, A. A., & Sumertajaya, I. M. (2011). Sidik Peubah Ganda dengan Menggunakan SAS. Dalam G. N. Wibawa, & A. F. Hadi. Bogor: IPB Press.
- Muhtadi. (2017). PENERAPAN PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM ALGORITMA K-MEANS UNTUK MENENTUKAN CENTROID PADA *CLUSTERING*. *Journal of Mathematic Teaching*.
- Nahdliyah, M. A., & dkk. (2019). METODE k-MEDOIDS *CLUSTERING* DENGAN VALIDASI SILHOUETTE INDEX DAN C-INDEX (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018). *JURNAL GAUSSIAN, Volume 8, Nomor 2, Tahun 2019*.
- Rasyid, A., Purba, S., & A, N. R. (2020). *Kaleidoskop 2020: Terbongkarnya Drama Perselingkuhan di Balik Pembunuhan Hakim PN Medan*. Diambil kembali dari sumut.inews.id: <https://sumut.inews.id/berita/kaleidoskop-2020-terbongkarnya-drama-perselingkuhan-di-balik-pembunuhan-hakim-pn-medan>
- Setyawati, A. W. (2017). *Implementasi Algoritma Partitioning Around Medoid (PAM) untuk Pengelompokan Sekolah Menengah Atas di DIY Berdasarkan Nilai Daya Serap Ujian Nasional*. Yogyakarta: Skripsi Fakultas Sains dan Teknologi Universitas Sanata Dharma.
- Simamora. (2005). *Analisis Multivariat Pemasaran*. Jakarta: PT. Gramedia Pustaka Utama.

- Supranto, J. (2004). *Analisis Multivariat : Arti dan Interpretasi*. Jakarta: PT. Rineka Cipta.
- Supratman, A. (2020). Menyoal Sikap Kejahatan Di Indonesia Di Era Industri 4.0 (Suatu Perspektif Kriminologi). *Leg. J. Perundang Undangan dan Huk. Pidana Islam, vol. 0*.
- Tampubolon, H. D., & dkk. (2021). Penerapan Algoritma K-Means dan K-Medoids *Clustering* untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi. *JURNAL ILMU KOMPUTER DAN TEKNOLOGI Vol 2 No 2*.
- Triaynto, W. A. (2015). *Algoritma K-Medoids untuk Penentuan Strategi Pemasaran Produk*. Jakarta: PT. Raja Grafindo Persada.
- Widarjono, A. (2009). *Ekonometrika Pengantar dan Aplikasinya*. Yogyakarta: Ekonisia.

LAMPIRAN

Lampiran 1 Data Kriminalitas di Indonesia Tahun 2020

Sumber: Badan Pusat Statistika

Kepolisian	X1	X2	X3	X4	X5	X6	X7	X8
Aceh	15	1206	232	29	145	2408	1596	1130
Sumatera Utara	99	6207	774	55	780	10916	5932	5562
Sumatera Barat	20	1724	294	97	185	3732	913	989
Riau	13	1013	143	11	413	3152	1646	1095
Jambi	11	531	69	9	87	1770	619	767
Sumatera Selatan	77	1551	172	141	563	2911	2554	1766
Bengkulu	33	479	108	1	106	963	340	556
Lampung	21	647	247	41	399	2532	1521	1373
Kepulauan Bangka Belitung	10	150	52	47	42	690	332	130
Kepulauan Riau	13	514	109	1	103	1003	362	543
DKI Jakarta	32	1449	113	31	259	3153	5981	4595
Jawa Barat	20	1445	371	11	257	2950	1141	2792

Jawa Tengah	30	382	349	11	236	3883	1601	1814
DI Yogyakarta	6	573	98	27	126	1769	460	1225
Jawa Timur	55	1404	256	177	341	4976	2629	2817
Banten	8	521	81	151	94	1657	634	718
Bali	6	414	65	4	54	821	668	408
Nusa Tenggara Barat	11	646	95	46	284	2120	571	876
Nusa Tenggara Timur	40	1726	160	14	42	1494	41	574
Kalimantan Barat	35	184	137	112	54	1254	755	431
Kalimantan Tengah	17	305	115	1	86	837	590	315
Kalimantan Selatan	38	462	51	0	367	1208	1582	539
Kalimantan Timur	12	352	113	1	151	862	1354	356
Kalimantan Utara	10	91	32	10	19	340	245	72
Sulawesi Utara	74	1976	242	88	72	1394	100	945
Sulawesi Tengah	18	1391	175	22	140	2121	340	677
Sulawesi Selatan	71	3848	172	311	348	4428	1013	1786
Sulawesi Tenggara	26	734	173	9	42	632	192	122
Gorontalo	13	1043	129	73	4	696	101	467

Sulawesi Barat	8	505	64	7	15	607	175	244
Maluku	10	993	1398	119	23	894	142	307
Maluku Utara	1	255	100	13	14	112	120	49
Papua Barat	10	655	47	124	155	1301	120	356
Papua	27	1296	100	176	532	3678	241	701

Lampiran 2 Sintaks

```
# Load Library

#`` `{r message=FALSE, warning=FALSE}
library(knitr)
install.packages("REdaS")
library(REdaS)
library(factoextra)
library(clValid)
library(tidyverse)
library(cluster)
#``

# Baca dataku

#`` {r}
dataku <- read.csv(file.choose(), header = T, sep = ";")
View(dataku)
boxplot(dataku[,2:9])
summary(dataku)

# Pengecekan asumsi

bart_spher(dataku[,2:9])
kmo(dataku[,2:9])

cor(dataku[,2:9])

#menghitung vektor eigen dan nilai eigen
R <- cor(dataku[,2:9])
eigen<- eigen(R)
eigen$values #eigen dapat 2

#Melakukan PCA
pcadata<- prcomp(x = dataku[,2:9], scale. = TRUE, center = TRUE)

#menentukan jumlah komponen utama
summary(pcadata)

#plot pca
plot(pcadata, type="lines")

#persamaan komponen utama
round(pcadata$rotation,2) #loading data
round(pcadata$sdev^2,2) #eigen

fviz_pca(pcadata) #Visualisasi data hasil rekonstruksi

#data pca fix
PCA_scoresbaru <- pcadata$x[,1:2]
View(PCA_scoresbaru)

new_Data=as.data.frame(PCA_scoresbaru)
new_Data
View(new_Data)
uji_bart(new_Data)
```

```

#k-medoids

library(tidyverse) # dataku manipulation
library(cluster) # clustering algorithms
library(factoextra) # clustering algorithms & visualization

# pamk (penentuan jumlah cluster)
library(fpc)
pamk.result <- pamk(PCA_scoresbaru)
pamk.result
pamk.result$nc

#menampilkan grafik sillhouette
fviz_nbclust(PCA_scoresbaru, pam, method = "silhouette")
pam.hasil <- pam(PCA_scoresbaru, 2)

#jarak
pam.hasil$diss

#datakuframe hasil cluster
df.clusterbaruuuu = data.frame(PCA_scoresbaru,pam.hasil$cluster)
View(df.clusterbaruuuu)

#Clustering
#sesuai abjad
table(pam.result$clustering, dataku$Kepolisian)

#plot cluster
fviz_cluster(pam.hasil, data = PCA_scoresbaru)

# Validasi Cluster
intern <- clValid(PCA_scoresbaru, 2:5, clMethods =
c("hierarchichal", "pam"), validation = "internal")
summary(intern)
optimalScores(intern)

# Dendrogram

#single linkage
data.hcc1 <- PCA_scoresbaru %>%
  dist(method = "euclidean") %>%
  hclust(method = "single")
fviz_dend(data.hcc1, k = 2,
          cex = 0.6, palette = "jco",
          rect = TRUE, rect_border = "jco", rect_fill = TRUE,
          main = "Cluster Dendrogram Single Linkage")

#complete linkage
data.hcc2 <- PCA_scoresbaru %>%
  dist(method = "euclidean") %>%
  hclust(method = "complete")
fviz_dend(data.hcc2, k = 2,
          cex = 0.6, palette = "jco",
          rect = TRUE, rect_border = "jco", rect_fill = TRUE,
          main = "Cluster Dendrogram Complete Linkage")

#average
data.hcc3 <- PCA_scoresbaru %>%
  dist(method = "euclidean") %>%

```

```

    hclust(method = "average")
fviz_dend(data.hcc3, k = 2,
          cex = 0.6, palette = "jco",
          rect = TRUE, rect_border = "jco", rect_fill = TRUE,
          main = "Cluster Dendrogram Average Linkage")

#ward
data.hcc4 <- PCA_scoresbaru %>%
  dist(method = "euclidean") %>%
  hclust(method = "ward.D2")
fviz_dend(data.hcc4, k = 2,
          cex = 0.6, palette = "jco",
          rect = TRUE, rect_border = "jco", rect_fill = TRUE,
          main = "Cluster Dendrogram (Ward's Method)")

#pemilihan metode terbaik
m <- c("single", "complete", "ward", "average")
names(m) <- c("single", "complete", "ward", "average")
ac <- function(x) {
  agnes(PCA_scoresbaru, method=x)$ac
}
map_dbl(m, ac)

# Deskripsi Statistik Cluster k-medoids
clust1 <- dataku[-c(2, 6, 11, 15, 27, 34),]
summary(clust1)
clust2 <- dataku[c(2, 6, 11, 15, 27, 34),]
summary(clust2)

# Deskripsi Statistik Cluster hierarchical
clust1_hieararki <- dataku[-c(22, 6, 11, 15, 27),]
summary(clust1)
clust2_hierarchical <- dataku[c(2, 6, 11, 15, 27),]
summary(clust2)

```