

*Cluster Analysis Using The K-Means Algorithm And Centroid Linkage
On 2021 Of Indonesian Crime Data*

**Analisis *Cluster* Menggunakan Algoritma *K-Means* dan
Centroid Linkage pada Data Kriminalitas Indonesia
Tahun 2021**

Adjie Dwi Sulistyono¹, Annisa Nur Hanifah²,
Isnandar Slamet³✉, Arisman Adnan⁴

^{1,2,3} Universitas Sebelas Maret, Surakarta, Indonesia

⁴ Universitas Riau, Pekanbaru, Indonesia

✉ isnandar slamet@staff.uns.ac.id

Received: 01-03-2024

Revised: 15-04-2024

Accepted: 30-04-2024

ABSTRACT

Crime is anything that violates the law or a crime. There are various types of crimes, such as crimes against life, physical, corruption, and so on. One method that can be used to analyze crime data is the cluster analysis method, a multivariate analysis technique that aims to cluster observational data or variables into groups with close proximity and a high degree of similarity. The purpose of the clustering is to find out areas that have similar types of crime and to reduce the number of crimes in Indonesia that often occur. In cluster analysis, there are two methods, namely hierarchical and non-hierarchical cluster methods. In this study, the Centroid Linkage and K-Means methods were used with a number of clusters 2. Based on cluster analysis, provinces in Indonesia can be grouped according to crime intensity in 2020. The average of each variable in the cluster members is used to determine the characteristics and name of the cluster.

Keywords: *crime; k-means; centroid linkage*

ABSTRAK

Kejahatan adalah segala sesuatu yang melanggar hukum atau bentuk criminal kejahatan. Terdapat beberapa jenis kejahatan, diantaranya kejahatan terhadap jiwa atau nyawa, kejahatan fisik, korupsi, dan lain sebagainya. Salah satu metode yang dapat digunakan untuk menganalisis data kejahatan adalah metode analisis *cluster*, yaitu suatu teknik analisis multivariat yang bertujuan untuk mengelompokkan data observasi atau variabel ke dalam kelompok-kelompok yang



mempunyai kedekatan dan tingkat kemiripan yang tinggi. Tujuan dari penelitian ini adalah untuk mengetahui daerah-daerah yang memiliki kesamaan jenis kejahatan dan untuk mengurangi jumlah kejahatan di Indonesia yang sering terjadi. Dalam analisis *cluster* terdapat dua metode yaitu metode *cluster* hierarki dan non-hierarki. Dalam penelitian ini digunakan metode *Centroid Linkage* dan *K-Means* dengan jumlah *cluster* 2. Berdasarkan analisis *cluster*, provinsi-provinsi di Indonesia dapat dikelompokkan berdasarkan intensitas kejahatan pada tahun 2020. Rata-rata dari setiap variabel anggota *cluster* digunakan untuk menentukan karakteristik dan nama *cluster*.

Kata kunci: kejahatan; k-means; centroid linkage

PENDAHULUAN

Kriminalitas adalah suatu tindakan yang merugikan orang lain, disamping itu tindakan ini ditetapkan sebagai bentuk pelanggaran undang-undang. Tindakan kriminalitas dapat dilakukan oleh semua orang, tanpa memandang tingkatan usia, jenis kelamin atau status sosial. Banyak faktor yang mempengaruhi tindak kejahatan, seperti faktor kesejahteraan, pendidikan, ekonomi, lingkungan hidup dan pekerjaan. Kriminalitas memang menjadi masalah sosial yang membuat semua orang resah, oleh karenanya diperlukan suatu informasi tempat kerawanan yang akurat untuk membantu mereka agar lebih berhati-hati. Informasi tindak kejahatan juga penting bagi kepolisian dalam rangka untuk mengantisipasi terjadinya tindak kejahatan.

Menurut Badan Pusat Statistik (BPS), jumlah kriminalitas atau kejahatan di Indonesia pada tahun 2020 mengalami penurunan dibandingkan tahun sebelumnya. Terdapat jumlah kejahatan pada tahun 2018 sebanyak 294281, pada tahun 2019 sebanyak 269324, dan pada tahun 2020 terdapat sebanyak 247218. Jumlah kejahatan tertinggi tercatat pada Kepolisian Daerah Sumatera Utara, yaitu sebanyak 32990 kejadian dan terendah berada pada Kepolisian Daerah Maluku Utara dengan 850 kejadian.¹

Jenis-jenis kejahatan secara umum terdiri atas 9 jenis kejahatan, yaitu kejahatan terhadap nyawa, fisik/badan, kesusilaan, kemerdekaan orang, hak milik/barang dengan penggunaan kekerasan, hak milik/barang tanpa penggunaan kekerasan, kejahatan terkait narkoba, kejahatan terkait penipuan, penggelapan, dan korupsi, serta kejahatan terhadap ketertiban umum.

¹Badan Pusat Statistik. "Statistik Kriminal 2021". <https://www.bps.go.id/publication//statistik-kriminal-2021/> (accessed : 17 Juni 2022).

Analisis *cluster* merupakan salah satu analisis pada *data mining*, yaitu *unsupervised* yang pada dasarnya digunakan mencari dan mengelompokkan data yang memiliki kemiripan karakteristik antara satu data dengan data yang lain. Dalam *data mining* ada dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering*. Salah satu metode pada *hierarchical clustering* yaitu *Centroid Linkage*.

Penelitian yang dilakukan Oyelade menggunakan 79 data mahasiswa untuk uji coba *clustering* pada Universitas Nigeria,² sedangkan penelitian Shovon & Haque menggunakan 60 data mahasiswa untuk uji coba penelitiannya.³ Penelitian menggunakan metode *Centroid Linkage* dilakukan oleh Silvi yang mengelompokkan provinsi berdasarkan 5 indikator faktor risiko prevalensi HIV/AIDS dengan jumlah 7 *cluster* dengan menggunakan kedua metode.⁴ Salah satu keuntungan metode *Centroid Linkage*, yaitu *outlier* tidak berpengaruh signifikan jika dibandingkan metode lain.

Berdasarkan beberapa penelitian yang telah disebutkan di atas, terlihat bahwa metode *K-Means* dan *Centroid Linkage* terbukti optimal untuk melakukan analisis *cluster*. Dari semua uraian permasalahan di atas, tindak kriminalitas atau kejahatan di Indonesia diteliti dengan menggunakan algoritma *K-Means* dan *Centroid Linkage*.

METODE PENELITIAN

Data yang digunakan pada penelitian ini merupakan data sekunder yang bersumber dari *website* Badan Pusat Statistik (BPS) yaitu publikasi Statistik Kriminal 2021. Data merupakan jumlah kriminalitas menurut jenisnya yang tercatat oleh Kepolisian Daerah setempat. Jumlah data yang digunakan sebanyak 34 provinsi yang ada di Indonesia. Variabel yang digunakan pada penelitian disajikan pada Tabel 1.

²Oyelade O.J. "Application Of K-Means Clustering Algorithm For Prediction Of Students' Academic Performance", International Journal Of Computer Science And Information Security, Vol. 7. 2010.

³Oyelade O.J. "Application Of K-Means Clustering Algorithm For Prediction Of Students' Academic Performance", International Journal Of Computer Science And Information Security, Vol. 7. 2010.

⁴Silvi R. "Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia". Jurnal Matematik "MANTIK", Vol.4(1): 22-31. 2018.

Tabel 1. Variabel Penelitian

| Variabel | Keterangan |
|------------------------|-----------------------------------------------------------------|
| <i>Key</i> Provinsi | <i>Attribute:</i> Nama-nama provinsi yang ada di Indonesia |
| X₁ | Kejahatan terhadap nyawa |
| X₂ | Kejahatan terhadap fisik |
| X₃ | Kejahatan terhadap kesusilaan |
| X₄ | Kejahatan terhadap kemerdekaan orang |
| X₅ | Kejahatan terhadap hak milik/barang dengan penggunaan kekerasan |
| X₆ | Kejahatan terhadap hak milik/barang tanpa penggunaan kekerasan |
| X₇ | Kejahatan terkait narkoba |
| X₈ | Kejahatan terkait penipuan, penggelapan, dan korupsi. |
| X₉ | Kejahatan terhadap ketertiban |

Langkah-langkah analisis yang dilakukan pada penelitian adalah sebagai berikut:

1. Mencari data dan studi literatur.
2. Memasukkan data ke dalam *software* Rstudio.
3. Melakukan analisis deskriptif.
4. Melakukan *preprocessing* pada data.
5. Melakukan analisis *clustering* menggunakan metode *K-Means*.
6. Melakukan analisis *clustering* menggunakan metode *Centroid Linkage*.
7. Menarik kesimpulan.

HASIL DAN PEMBAHASAN

Kriminalitas

Kriminalitas atau kejahatan memang merupakan masalah yang sangat umum yang terjadi di masyarakat di manapun berada, tindakan kriminalitas dapat terjadi di beberapa tempat dengan jangka waktu yang sama atau berbeda. Kriminalitas adalah suatu perbuatan yang dapat menimbulkan masalah-masalah dan keresahan bagi kehidupan di dalam Masyarakat.⁵ Kejahatan adalah yang memiliki dua macam pengertiannya, yaitu secara yuridis dan secara sosiologi. Secara yuridis formal, kejahatan adalah tingkah laku kejahatan yang melanggar

⁵Friedrichs, D. O. "Crimes of the powerful and the definition of crime". In *The Routledge international handbook of the crimes of the powerful* (pp. 39-49). Routledge. 2015.

hukum pidana yang ada. Pengertian secara sosiologi adalah meliputi segala tingkah laku manusia, walaupun tidak atau belumnya ditentukan dengan undang-undang.⁶

Data Mining

Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode *data mining* ini dapat digunakan untuk mengambil keputusan di masa depan. *Data mining* adalah pengolahan data dengan skala besar, sehingga *data mining* memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi.⁷ *Data mining* juga dapat dilakukan pada berbagai jenis *database* dan penyimpanan informasi, tetapi jenis pola yang akan ditemukan ditentukan oleh berbagai fungsi *data mining* seperti deskripsi *class*/konsep, asosiasi, analisis korelasi, klasifikasi, prediksi, analisis *cluster* dan lain-lain.

Clustering

Pada dasarnya *clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan dengan tidak memerlukan target *output*. Dalam *data mining* ada dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering*.⁸

Metode *hierarchical clustering* adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut. Metode *non-hierarchical clustering* justru dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan. Setelah jumlah *cluster* diketahui, baru proses *cluster* dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan *K-Means Clustering*.⁹

⁶Downes, D.M., Rock, P.E. and McLaughlin, E. "Understanding deviance: A guide to the sociology of crime and rule-breaking". Oxford University Press. 2016.

⁷ Han, J., Pei, J. and Kamber, M. "Data mining: concepts and techniques". Elsevier. 2011.

⁸ Rokach, L. and Maimon, O. "Clustering methods". In Data mining and knowledge discovery handbook (pp. 321-352). Springer, Boston, MA. 2005.

⁹ Milligan, G.W. and Cooper, M.C. "Methodology review: Clustering methods". Applied psychological measurement, 11(4), pp.329-354. 1987.

K-Means

K-Means merupakan salah satu metode *clustering non-hirarki* yang berusaha mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

Untuk tahapan-tahapan pengelompokkan menggunakan metode *K-Means* dijelaskan sebagai berikut⁹:

- Menentukan banyaknya k jumlah *cluster* yang digunakan.
- Membangkitkan *centroid* secara acak dengan persamaan sebagai berikut;
- Menghitung jarak *centroid* menggunakan *Euclidean Distance* dengan persamaan sebagai berikut:

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

dimana:

$D(i, j)$: jarak data ke- i ke pusat *cluster* j

X_{ki} : data ke- i pada atribut data ke- k

X_{kj} : titik pusat ke- j pada atribut ke- k

- Mengelompokkan setiap observasi berdasarkan jarak terdekat dengan *centroid*-nya.
- Menentukan posisi *centroid* baru dengan cara menghitung rata-rata dari data yang berbeda dengan *centroid* yang sama.

Hierarchical Clustering

Tipe dasar dalam metode *hierarchical clustering* bisa *agglomerative* atau *devisif*. Ada lima metode *hierarchical clustering agglomerative* dalam pembentukan *cluster* yaitu *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Ward's Linkage*, dan *Centroid Linkage*.¹⁰ Pada penelitian ini akan digunakan salah satu metodenya yaitu *Centroid Linkage*. *Centroid* adalah rata-rata semua objek dalam *cluster*. Pada metode *centroid* ini, jarak antar *cluster* adalah jarak antar *centroid*. *Centroid* baru dihitung ketika setiap kali obyek digabungkan, sehingga setiap kali anggotanya bertambah maka *centroid*nya akan berubah. Jarak antar *cluster* pada metode ini adalah jarak antar *centroid*.¹¹

¹⁰Zhao, Y., Karypis, G. and Fayyad, U. "Hierarchical clustering algorithms for document datasets". Data mining and knowledge discovery. Vol.10(2), pp.141-168. 2005.

¹¹Johnson R A dan Wichern D W. "Applied Multivariate Statistical Analysis Third Edition". New Jersey: Prentice Hall International. 1992.

Untuk langkah-langkah pengelompokkan menggunakan metode *Centroid Linkage* dijelaskan sebagai berikut:¹²

- Membuat k cluster, dimana masing-masing memuat satu kesatuan. Kemudian dibuat matrik jaraknya (dari i ke kelompoknya), dengan rumus: $D = \{d_{ik}\}$.
- Mencari matriks jarak untuk pasangan *cluster* terdekat, yaitu d_{uv} (jarak cluster u dan cluster v).
- Menggabungkan *cluster* u dan v menjadi *cluster* baru w , kemudian perbaharui matriks jaraknya.
- Ulangi langkah b & c sebanyak $k-1$ kali. Catat identitas dan tingkat (jarak atau similaritas) dari cluster yang digabungkan.

Sebelum melakukan penelitian, dilakukan studi literatur terlebih dahulu mengenai jenis-jenis kriminalitas/kejahatan yang terjadi di Indonesia melalui publikasi BPS tahun 2021. Dengan demikian, penelitian membahas sembilan jenis kejahatan yang terjadi di Indonesia.

Analisis Deskriptif

Untuk mengetahui karakteristik data yang digunakan dalam penelitian ini dapat dilihat dari statistik deskriptif yang disajikan pada Tabel 2.

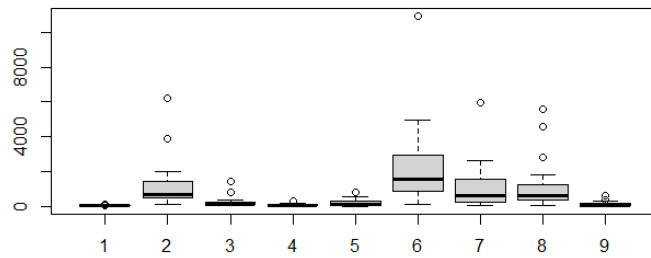
Tabel 2. Deskripsi Intensitas Kriminalitas di Indonesia

| Variabel | Min | Maks | Median | Mean |
|----------|-----|-------|--------|----------|
| X1 | 1 | 99 | 19 | 26.41176 |
| X2 | 91 | 6207 | 656 | 1078.882 |
| X3 | 32 | 1398 | 122 | 202.1176 |
| X4 | 0 | 311 | 28 | 57.94118 |
| X5 | 4 | 780 | 133 | 192.2941 |
| X6 | 112 | 10916 | 1575.5 | 2154.824 |
| X7 | 41 | 5981 | 604.5 | 1076.794 |
| X8 | 49 | 5562 | 625.5 | 1073.794 |
| X9 | 2 | 632 | 45.5 | 107.7059 |

¹²Zhao, Y., Karypis, G. and Fayyad, U. "Hierarchical clustering algorithms for document datasets". Data mining and knowledge discovery. Vol.10(2), pp.141-168. 2005.

Preprocessing

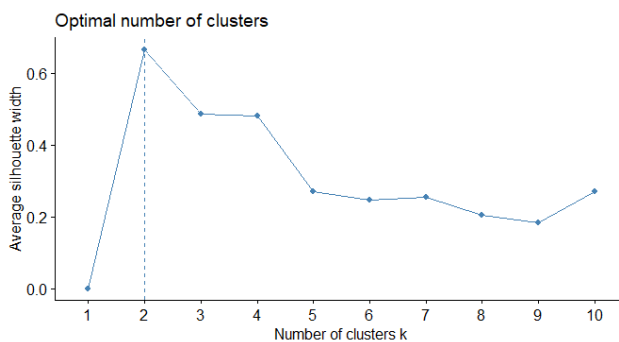
Sebelum masuk tahap analisis *clustering*, data dilakukan *preprocessing* terlebih dahulu. Tahap awal adalah pengecekan *missing value* pada *dataset* awal. Setelah dicek dapat diketahui bahwa tidak terdapat *missing value*. Setelah itu dilakukan pengecekan *outlier* dan didapatkan terdapat *outlier* pada semua variabel seperti Gambar 1. Karena data merupakan jumlah kriminalitas dan berjumlah sedikit maka *outlier* ini diabaikan.



Gambar 1. Boxplot Data

K-Means

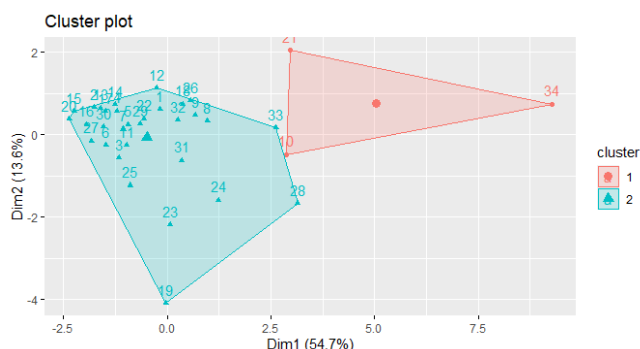
Pada metode *K-Means* terdapat beberapa pertimbangan untuk menentukan jumlah *cluster* yang paling optimal. Pendekatan metode yang paling populer di antaranya adalah metode *Silhouette* yang menggunakan nilai rata-rata *Silhouette*, dengan asumsi semakin tinggi nilai rata-ratanya maka k tersebut semakin optimal. Berikut disajikan *output* metode *Silhouette* pada Gambar 2 dengan menggunakan *software* RStudio.



Gambar 2. Output R Metode *Silhouette*

Penentuan Anggota Cluster. Penentuan *cluster* dapat dilakukan dengan berbagai metode, salah satunya dengan metode *K-Means*. Algoritma *K-Means* melakukan klusterisasi berbasis titik dengan menentukan tiga parameter, yaitu jumlah *cluster*, inialisasi *cluster*, dan jarak sistem. Hasil

analisis yang didapatkan dengan klaterisasi menggunakan *K-Means* seperti Gambar 3.



Gambar 3. Cluster Plot *K-Means*

Berdasarkan Gambar 3, didapatkan bahwa *cluster 1* terdiri dari tiga provinsi, sedangkan 33 provinsi lainnya masuk dalam *cluster 2*. Adapun nama-nama provinsi berdasarkan clusterisasi sebagaimana Tabel 3 berikut.

Tabel 3. Anggota *Cluster* Metode *K-Means*

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

Karakteristik *Cluster*. Karakteristik *cluster* pada penelitian ini dilakukan berdasarkan rata-rata intensitas kejahatan pada masing-masing *cluster* yang disajikan pada

Tabel 4 yang selanjutnya dapat disimpulkan menjadi nama *cluster*.

Tabel 4. Karakteristik *Cluster* Metode *K-Means*

| Jenis Kekerasan | <i>Cluster</i> | |
|-----------------|----------------|----|
| | 1 | 2 |
| Nyawa | 62 | 23 |

| | | |
|---------------------------------------|------|------|
| Fisik | 3020 | 891 |
| Kesusilaan | 381 | 185 |
| Kemerdekaan Orang | 87,7 | 55,1 |
| Hak Milik dengan Kekerasan | 460 | 166 |
| Hak Milik Tanpa Kekerasan | 6348 | 1749 |
| Narkotika | 4847 | 712 |
| Penipuan, Penggelapan, dan Korupsi | 4325 | 759 |
| Ketertiban | 225 | 96,4 |

Berdasarkan

Tabel 4, didapatkan bahwa *cluster 1* memiliki tingkat kriminalitas yang lebih tinggi dibandingkan dengan *cluster 2*, meliputi kekerasan pada nyawa, fisik, kesusilaan, kemerdekaan orang, hak milik dengan kekerasan, hak milik tanpa kekerasan, narkotika, serta penipuan, penggelapan, dan korupsi. Sedangkan, pada *cluster 1*, kriminalitas dalam hal ketertiban memiliki tingkat yang rendah jika dibandingkan *cluster 2*. Persebaran daerah *cluster* rawan dan tidak rawan kriminalitas disajikan secara geografis pada Gambar 4 berikut.

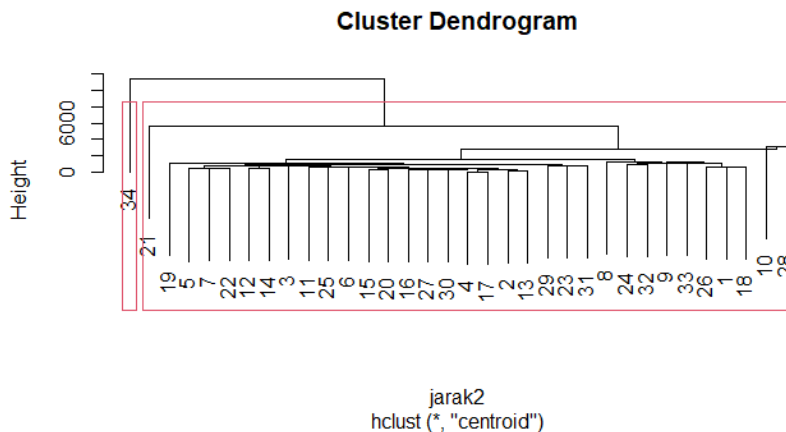


Gambar 4. Persebaran Kriminalitas dengan *K-Means*

Centroid Linkage

Salah satu metode pada *hierarchical clustering* yaitu *Centroid Linkage* yang pertimbangan jarak antar *cluster* dengan jarak *centroid*.

Penentuan Anggota Cluster. Pada metode ini, kami juga menggunakan jumlah *cluster* k sebanyak 2. Penentuan anggota *cluster* disajikan dalam dendogram pada Gambar 5 berikut.



Gambar 5. Dendrogram *Centroid Linkage*

Berdasarkan Gambar 5, didapatkan bahwa terdapat 33 provinsi yang masuk dalam kategori cluster 1 dan satu provinsi yang masuk dalam kategori cluster 2. Adapun nama-nama provinsi berdasarkan clusterisasi sebagaimana Tabel 5 berikut.

Tabel 5. Anggota *Cluster* Metode *Centroid Linkage*

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

Karakteristik Cluster. Karakteristik *cluster* pada penelitian ini dilakukan berdasarkan rata-rata intensitas kejahatan pada masing-masing *cluster* yang disajikan pada

Tabel 6 yang selanjutnya dapat disimpulkan menjadi nama *cluster*.

Tabel 6. Karakteristik *Cluster* Metode *Centroid Linkage*

| Jenis Kekerasan | Cluster | |
|------------------------------------|---------|--------|
| | 1 | 2 |
| Nyawa | 242 | 99 |
| Fisik | 923 | 6.207 |
| Kesusilaan | 185 | 774 |
| Kemerdekaan Orang | 58 | 55 |
| Hak Milik dengan Kekerasan | 174 | 780 |
| Hak Milik Tanpa Kekerasan | 1.889 | 10.916 |
| Narkotika | 930 | 5.932 |
| Penipuan, Penggelapan, dan Korupsi | 938 | 5.562 |
| Ketertiban | 103 | 249 |

Berdasarkan

Tabel 6, didapatkan bahwa *cluster* 1 memiliki tingkat kriminalitas yang lebih rendah dibandingkan dengan *cluster* 2, meliputi kekerasan pada nyawa, fisik, kesusilaan, hak milik dengan kekerasan, hak miliki tanpa kekerasan, narkotika, penipuan, penggelapan, dan korupsi, serta ketertiban. Sedangkan, pada *cluster* 1, kriminalitas dalam hal kemerdekaan orang memiliki tingkat yang rendah jika dibandingkan *cluster* 2. Persebaran daerah *cluster* rawan dan tidak rawan kriminalitas disajikan secara geografis pada Gambar 6 berikut.



Gambar 6. Persebaran Kriminalitas dengan *Centroid Linkage*

KESIMPULAN

Berdasarkan hasil pembahasan, dapat disimpulkan berbagai hal berikut :

1. Analisis *clustering* menggunakan metode *K-Means* didapatkan jumlah *cluster* optimal sebanyak 2 *cluster*. *Cluster* 1 merupakan *cluster* dengan intensitas kriminalitas yang lebih tinggi daripada *cluster* 2 pada semua jenis kriminalitas.
2. Anggota dalam *cluster* 1 merupakan daerah yang rawan kriminalitas. Terdapat sebanyak tiga Provinsi yang termasuk didalamnya yaitu Provinsi Jawa Timur, Metro Jaya, dan Sumatera Utara. Sedangkan 31 Provinsi lainnya masuk ke dalam *cluster* 2 yang merupakan daerah tidak rawan kriminalitas. Daerah yang termasuk di dalamnya, yaitu Provinsi Aceh, Bali, Banten, Bengkulu, DI Yogyakarta, Gorontalo, Jambi, Jawa Barat, Jawa Tengah, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kep. Bangka Belitung, Kep. Riau, Lampung, Maluku, Maluku Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua, Papua Barat, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Barat, dan Sumatera Selatan.
3. Analisis *clustering* dengan metode *Centroid Linkage* digunakan k sebanyak 2. *Cluster* 1 merupakan *cluster* dengan intensitas kriminalitas yang lebih rendah daripada *cluster* 2 pada semua jenis kriminalitas, kecuali kejahatan terhadap kemerdekaan orang.

Anggota dalam *cluster* 1 merupakan daerah yang tidak rawan kriminalitas. Terdapat sebanyak 33 Provinsi yang termasuk di dalamnya, yaitu Provinsi Aceh, Bali, Banten, Bengkulu, DI Yogyakarta, Gorontalo, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kep. Bangka Belitung, Kep. Riau, Lampung, Maluku, Maluku Utara, Metro Jaya, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua, Papua Barat, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Barat, dan Sumatera Selatan. Sedangkan satu Provinsi lainnya masuk ke dalam *cluster* 2 yang merupakan daerah rawan kriminalitas, yaitu Provinsi Sumatera Utara.

REFERENSI

- Badan Pusat Statistik. “Statistik Kriminal 2021”.
<https://www.bps.go.id/publication//statistik-kriminal-2021/> (accessed : 17 Juni 2022).
- Oyelade O J. “Application Of K-Means Clustering Algorithm For Prediction Of Students' Academic Performance”, International Journal Of Computer Science And Information Security, Vol. 7. 2010.

- Shovon M, dan Haque M. "An Approach Of Improving Student Academic Performance By Using K-Means Clustering Algorithm And Decision Tree". Vol.(3)8. 2012.
- Silvi R. "Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia". Jurnal Matematik "MANTIK", Vol.4(1): 22-31. 2018.
- Friedrichs, D. O. "Crimes of the powerful and the definition of crime". In *The Routledge International Handbook of the Crimes of the Powerful* (pp. 39-49). Routledge. 2015.
- Downes, D.M., Rock, P.E. and McLaughlin, E. "Understanding deviance: A guide to the sociology of crime and rule-breaking". Oxford University Press. 2016.
- Han, J., Pei, J. and Kamber, M. "Data mining: concepts and techniques". Elsevier. 2011.
- Rokach, L. and Maimon, O. "Clustering methods". In *Data mining and knowledge discovery handbook* (pp. 321-352). Springer, Boston, MA. 2005.
- Milligan, G.W. and Cooper, M.C. "Methodology review: Clustering methods". *Applied psychological measurement*, 11(4), pp.329-354. 1987.
- Zhao, Y., Karypis, G. and Fayyad, U. "Hierarchical clustering algorithms for document datasets". *Data mining and knowledge discovery*. Vol.10(2), pp.141-168. 2005.
- Johnson R A dan Wichern D W. "Applied Multivariate Statistical Analysis Third Edition". New Jersey: Prentice Hall International. 1992.
- Zhao, Y., Karypis, G. and Fayyad, U. "Hierarchical clustering algorithms for document datasets". *Data mining and knowledge discovery*. Vol.10(2), pp.141-168. 2005.